



Universitat Autònoma
de Barcelona

Departament d'Arquitectura de
Computadors i Sistemes Operatius

Màster en
Ciència i Enginyeria Computacional

Solució paral·lelitzada d'interpolació *kriging* amb ajust automatitzat del variograma

Memòria del treball de recerca del
“Màster en Ciència i Enginyeria
Computacional”, realitzada per LLUÍS
PESQUER MAYOS, sota la direcció d’
ANA CORTÉS FITÉ
Presentada a la Escola Tècnica Superior
d’ Enginyeria (Departament d’
Arquitectura de Computadors i
Sistemes Operatius)

2008

Treball de recerca
Màster en Ciència i Enginyeria Computacional
Curs 2007-08

Solució paral·lelitzada d'interpolació *kriging* amb ajust automatitzat del variograma

Autor: Lluís Pesquer Mayos
Director: Ana Cortés Fité
Departament d' Arquitectura de Computadors i Sistemes Operatius
Escola Tècnica Superior d' Enginyeria (ETSE)
Universitat Autònoma de Barcelona

Signat

Lluís Pesquer Mayos

Ana Cortés Fité

Resum

El principal objectiu d'aquest treball és proporcionar una metodologia per a reduir el temps de càlcul del mètode d'interpolació *kriging* sense pèrdua de la qualitat del model resultat. La solució adoptada ha estat la paral·lelització de l'algorisme mitjançant MPI sobre llenguatge C.

Prèviament ha estat necessari automatitzar l'ajust del variograma que millor s'adapta a la distribució espacial de la variable d'estudi. Els resultats experimentals demostren la validesa de la solució implementada, en reduir de forma significativa els temps d'execució final de tot el procés.

Resumen

El principal objetivo de este trabajo es proporcionar una metodología para reducir el tiempo de cálculo del método de interpolación *kriging* sin pérdida de la calidad del modelo resultado. La solución adoptada ha sido la paralelización del algoritmo mediante MPI sobre lenguaje C.

Previamente ha sido necesario automatizar el ajuste del variograma que mejor se adapta a la distribución espacial de la variable de estudio. Los resultados experimentales demuestran la validez de la solución implementada, al reducir de forma significativa los tiempos de ejecución finales del proceso completo.

Abstract

The main objective of this work is to provide a methodology to reduce the time needed to calculate the kriging interpolation method without losing any quality of the resulting model. The solution adopted has been the algorithm parallelization by MPI on C language.

Previously, it has been necessary to automate the variogram fitting that best suits the spatial distribution of the variable at study. The experimental results demonstrate the validity of the implemented solution, to significantly reduce the execution times of the entire process

Agraïments

Haig d'agrair el suport i la col·laboració, fonamental per assimilar aquells procediments que eren nous per a mi, de qualsevol professor del CAOS que he consultat, tothom ha estat molt disposat a donar un cop de mà i de forma molt significativa a l'Anna i l'Andrés.

Una gran part dels meus coneixements previs, que he procurat d'explicar en aquest treball, són fruit d'uns quants anys treballant al CREAF, però fonamentalment m'han contagiats la motivació per la recerca en un embolcall de gent molt especial. Els meus companys GRUMETS els considero també part d'aquest col·lectiu màgic. És injust citar només alguns noms, però no me'n puc estar de donar gràcies de forma molt especial al Joan i al Xavier, però també haig d'afegir que sense companys com l'Angi, la Ivette, en Pino, el Gerard, l'Arnald, els futboleros ... i molts més (insisteixo que és injust concretar noms) treballar al CREAF seria una altra cosa, i possiblement, en un ambient diferent no m'hagués animat a fer aquest treball. Gràcies.

Vull agrair al David del Jardí Botànic de Barcelona, la seva col·laboració en la cessió de dades, que ha resultat ser molt útil en els resultats experimentals.

Per últim, però en el lloc més destacat de tots, haig d'agrair el suport que m'ha donat la Maribel, que ha patit els efectes col·laterals d'aquest treball (com també el Martí i la Laia) i ha continuat recolzant-me en tot moment. Sense el seu esforç hagués estat impossible. GRÀCIES.

Índex

Prefaci	1
1. Interpolació espacial.....	3
1.1. Fonaments	3
1.2. Classificació dels mètodes d'interpolació	5
1.3. Mètodes d'interpolació representatius	7
1.3.1. Superfícies de tendència	7
1.3.2. Invers ponderat de la distància	7
1.3.3. <i>Splines</i>	8
1.3.4. <i>Kriging</i>	9
2. <i>Kriging</i>.....	11
2.1. Introducció	11
2.2. Geostatística	12
2.3. Variograma.....	14
2.4. Variants de <i>kriging</i>	16
2.5. <i>Kriging</i> ordinari.....	18
3. Ajust automatitzat del variograma	23
3.1. Introducció	23
3.2. Ajust interactiu del variograma.....	23
3.2.1. Construcció del variograma empíric	24
3.2.2. Generació del variograma modelitzat	25
3.3. Solució automatitzada per l'ajust del variograma	27
4. Paral·lelització <i>del kriging</i>	31
4.1. Justificació	31
4.2. Solució aproximada per reducció local.....	32
4.3. Solució completa per paral·lelització.....	34

4.3.1. Solució MPI.....	34
4.3.2. Disseny de la paral·lelització.....	
5. Resultats experimentals	39
5.1. Descripció de l'entorn experimental.....	39
5.2. Interpolació de temperatures.....	40
5.3. Interpolació d'altituds	45
6. Conclusions i treballs futurs	49
7. Bibliografia	51
Apèndix	57

Prefaci

En una altra cultura on el mètode científic estigués construït sobre uns fonaments diferents als de la ciència actual, la interpolació no existiria com a mètode per a predir situacions no observades o anticipar nous escenaris. En aquesta altra cultura, el paper que en el mètode científic té la interpolació, el poden representar els mètodes d'endevinació que es fonamenten en determinades intuïcions, creences, experiències paral·leles, etc.

Aquest treball es presenta en el marc de la Ciència Computacional, i per tant ha d'assumir (i les assumeix) les bases del mètode científic; però malgrat aquests sòlids fonaments, en la interpolació resta un substrat d'endevinació, que caldrà que sigui el més prim possible.

Precisament una de les característiques principals del mètode desenvolupat en aquest treball, el *kriging*, assumeix, a diferència d'altres alternatives deterministes, que els escenaris naturals són complexos i que en qualsevol modelització existirà una component de difícil explicació. Aquesta component en el llenguatge de la geostatística, que s'exposa en el desenvolupament del treball, s'anomena aleatòria i cal analitzar-la amb metodologies que d'entrada reconguin les seves limitacions predictives.

Modelitzar aquests escenaris complexos requereixen solucions més elaborades que algunes solucions deterministes, però aquesta complexitat implica processos amb elevades necessitats de càlcul i algunes prediccions són inviables si la resposta que proporciona el model és posterior a les accions que cal realitzar una vegada coneguda la predicció. Aquesta reflexió ha portat a l'autor a creure necessari implementar una solució que permeti al model elaborar una resposta ràpida mantenint la complexitat, el rigor i la qualitat del resultat, i l'autor ha cregut que la Ciència Computacional oferia una solució que resol aquesta problemàtica.

El treball que es presenta és un sinuós itinerari per diverses matèries d'anàlisi espacial i modelització numèrica abordades fonamentalment des de la Geografia i l'Estadística que necessiten dels instruments de la Ciència Computacional per a optimitzar la seves solucions. Els àmbits d'aplicació dels mètodes d'interpolació és molt extens i pot implicar a qualsevol matèria on la comprensió, l'anàlisi i la gestió del territori (o també espais més abstractes) constitueixin elements del seu estudi.

Amb aquest objectiu, i abans d'arribar a la solució de paral·lelització que ofereix la Ciència Computacional, es contextualitza el *kriging* dins els mètodes d'interpolació, s'exposen les bases de la interpolació geostatística i es descriu amb detall l'algorisme base d'aquest treball. Abans de passar a la pròpia implementació, es justifica la necessitat de l'automatització d'una de les parts més complexes del *kriging*, la modelització del variograma, i es mostra la solució adoptada.

Dins de les pròpies metodologies computacionals, el treball detalla els aspectes pràctics de la implementació, fent especial èmfasi en el disseny de la paral·lelització, resultat de l'anàlisi de distribució de càrrega de les diferents etapes del procés i de les condicions i paràmetres que poden participar.

Finalment es mostren els resultats de les proves realitzades en un entorn distribuït dels laboratoris del Departament d'Arquitectura de Computadors i Sistemes Operatius. Aquests resultats permeten assolir els objectius fixats abans d'iniciar el treball i confirmen la utilitat de les metodologies instrumentals que ofereix la Ciència Computacional a altres àmbits de la Ciència. També pot servir per justificar una visió personal molt poc compartimentada sobre el coneixement científic, una visió multidisciplinària que recordi a èpoques renaixentistes, que en el món actual es podria reetiquetar com el mestissatge de la Ciència.

Capítol 1

Interpolació espacial

1.1 Fonaments

D'una forma primària es pot definir la interpolació espacial com a un procés de transformació d'un model de dades puntual a un model de valors continu en el corresponent àmbit d'aplicació.

Per tant, segons aquesta definició, constituïria un procés d'ampliació del domini d'una variable, d'un domini reduït amb només unes quantes localitzacions a un domini extens a tot l'àmbit d'una determinada regió. Ara bé, aquesta definició pot donar lloc a interpretacions excessivament simplifiades, que situarien a la interpolació al mateix conjunt de metodologies de conversions entre models i formats de dades espacials.

Sense oblidar que la interpolació també té aquesta propietat, cal una definició que completi amb major precisió les característiques essencials i diferencials de la interpolació.

Una definició més elaborada, inspirada en la donada per [BUR98] que recull alguns dels seus trets més essencials pot ser:

Interpolació espacial és el procediment matemàtic que té per objectiu predir el valor incògnita d'un atribut en una localització precisa a partir de valors de l'atribut coneguts en localitzacions veïnes situades a l'interior de la mateixa regió d'estudi.

Implícits en aquesta definició, és important destacar dos aspectes:

- Quan la predicció implica també a zones exteriors a l'àmbit de les dades de mostra, ja no és adequat el concepte d'interpolació i cal referir-se a extrapolació, [LAU95].
- La interpolació espacial s'inscriu dins un domini espacial que ha d'estar referit en un determinat sistema de coordenades (CRS)

[ISO01], que en principi pot ser n-dimensional, però que dins de l'àmbit d'estudi d'aquest treball es restringirà a un espai planimètric de dues dimensions (2D), georeferenciat en base un sistema de referència horitzontal (SRH) [FDG98]. Això vol dir que no participaran espais amb la component altimètrica com a coordenada (podria ser-ho com a variable) ni espais d'anàlisi d'evolució temporal amb la coordenada temps afegida.

En conseqüència vol dir que es disposa d'una sèrie d'observables d'un atribut o variable en unes localitzacions concretes, exemple de la figura 1.1 , i es pretén elaborar un solució que permeti estimar la variable o atribut en qualsevol localització de tot un àmbit espacial interior a l'àmbit global del conjunt d'observables. Aquesta solució pot arribar a presentar-se per una fórmula analítica en alguns casos (p.ex. superfícies de tendència), un llistat de localitzacions problema o en la seva forma més comú, donar lloc a un model digital del terreny (MDT), o sigui un conjunt de dades numèriques que descriuen la distribució espacial d'una característica del territori [FEL94], exemple corresponent a la figura 1.2.

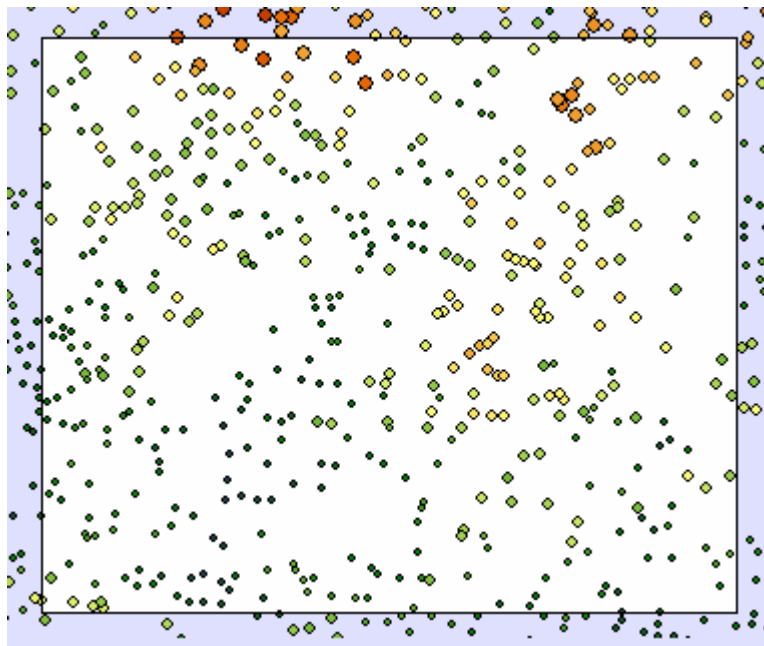


Figura 1.1: Il·lustració esquemàtica de la distribució de les localitzacions de possibles observables puntuals. En un marc, l'àmbit d'interpolació i en blau la regió externa.

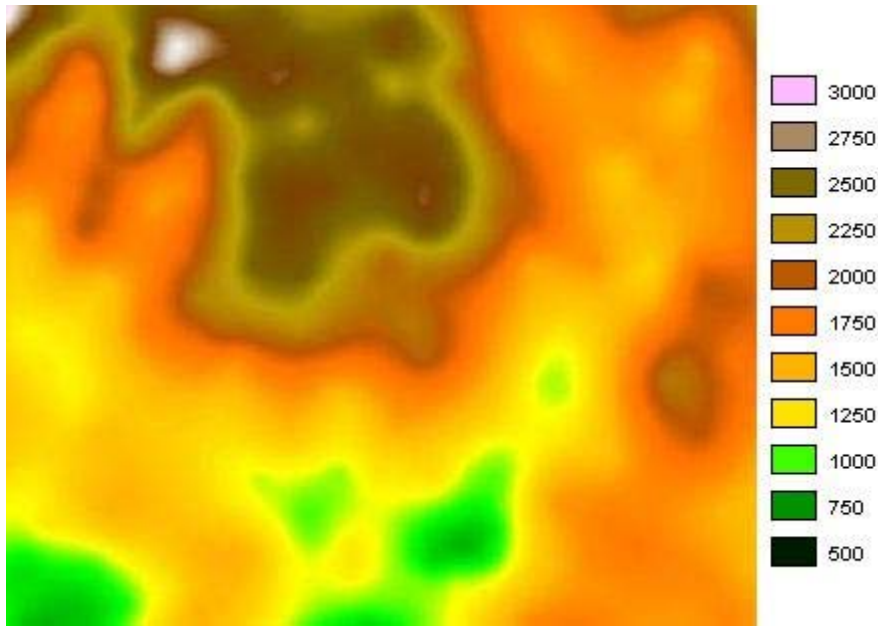


Figura 1.2: Mapa corresponent a un possible MDT, com a model ràster en una visualització per assignació continua entre valors i colors.

Cal remarcar que la hipòtesi bàsica de la interpolació és l'anàlisi de la variabilitat de l'atribut en relació a cada localització, absoluta en alguns casos i molt sovint relativa a la resta d'observacions. Això vol dir que, habitualment, valors d'un atribut dins d'un veïnatge espacial proper tenen una elevada probabilitat de ser similars i aquesta probabilitat disminueix respecte a valors a veïnatge allunyat. De fet, en el seu substrat habita la primera llei de la Geografia:

Everything is related to everything else, but near things are more related than distant things. [TOB70]

Alguns mètodes d'interpolació espacial també permeten modelitzar altres patrons espacials diferents d'aquest més comú, però són patrons poc freqüents en estudis de sistemes naturals, sistemes dels quals s'analitzarà en aquest treball, aplicacions concretes.

1.2 Classificació dels mètodes d'interpolació

Tradicionalment s'han classificat els mètodes d'interpolació en dos tipus:

- **Mètodes globals:** La solució generada a tot l'àmbit és única, en el sentit que, en tots els punts problema, o sigui en totes les localitzacions que es desitja conèixer el valor predit pel mètode d'interpolació, participen totes les dades observables alhora.
- **Mètodes locals:** Es genera un conjunt de solucions, cadascuna només hàbil dins d'un veïnatge proper al voltant de cada ubicació del

punt problema, o en alguns casos per a un conjunt de subregions de l'àmbit global.

La principal avantatge del mètode global és la seva homogeneïtat, i per tant la seva continuïtat com a funció matemàtica i de les seves derivades. S'usen fonamentalment per a analitzar patrons principals i establir tendències generals, però sovint els manca el nivell de detall suficient per a explicar amb precisió comportaments locals.

Per oposició, la principal virtut dels mètodes locals són la seva adequació a escales de detall, però a canvi d'introduir discontinuïtats en diferents entorns i no considerar el problema de forma conjunta.

Complementant aquestes dues tipologies extremes caldria afegir dos tipus intermedis, usats a la pràctica en moltes aplicacions com a solució equilibrada que pretén sumar les avantatges d'ambdues i minimitzar els seus inconvenients:

- **Mètodes quasi-locales amb propòsits globals:** Són essencialment mètodes locals que afegeixen condicions de contorn per a forçar una mínima continuïtat entre les diverses solucions adaptades localment.
- **Mètodes globals amb solucions regionalitzades:** Es construeixen a partir de la divisió en subregions de l'àmbit total on s'apliquen diverses solucions 'globals'. Si la subdivisió de la regió total s'elabora amb regions de solapament, la transició entre les diverses solucions pot donar lloc a models amb discontinuïtat reduïda com en l'exemple de la figura 1.3.

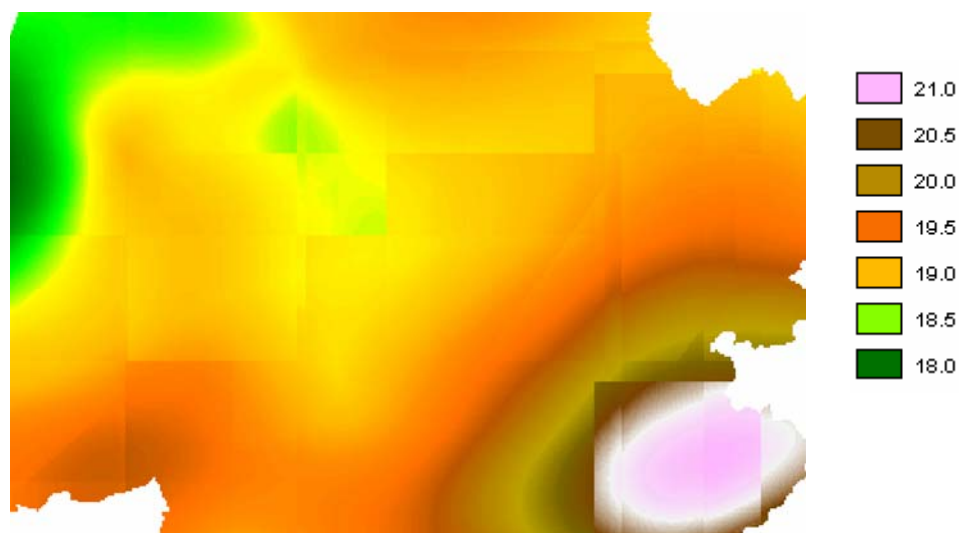


Figura 1.3: Il·lustració d'un mètode global subdividit en regions regulars on s'observen les discontinuïtats en les zones de transició.

En funció de la reproducció fidel o no, del valor observable per part del model interpolat en les pròpies localitzacions es classifiquen en dos tipus:

- **Interpoladors exactes:** el valor interpolat coincideix exactament amb el valor observat a cada punt de la mostra.

- **Interpoladors no exactes:** el model no reproduïx necessàriament el valor de cada mostra, però en conjunt es minimitzen les diferències (models de regressió). En determinats casos s'estableixen i apliquen límits màxims que limiten la diferència entre el valor interpolat i el valor observat.

1.3 Mètodes d'interpolació representatius

És extensa la bibliografia que analitza i compara els diversos mètodes d'interpolació espacial; [GEO94] i [FRA82] són dos estudis exhaustius on es poden trobar més de 30 mètodes diferents. Ara bé, la llista es redueix considerablement en referències de propòsit més ampli i menys específic com [BUR98], [BON94] i [OSU02] o en determinades aplicacions molt especialitzades [KRA06]. Aquestes opcions més representatives acostumen a coincidir amb les solucions implementades de forma majoritària en els diferents *softwares* d'anàlisi espacial o de Sistemes d'informació Geogràfica (SIG o GIS) [COO04]

A continuació es descriuen les característiques bàsiques dels mètodes considerats més representatius amb l'objectiu de contextualitzar el *kriging* dins el conjunt de mètodes d'interpolació i valorar amb criteris comparatius adequats les seves principals característiques.

1.3.1 Superfícies de tendència

Mètode d'interpolació global no-exacte que correspon a la determinació d'una funció polinòmica bidimensional de grau habitualment baix:

Els paràmetres que fixen una funció polinòmica concreta es determinen per un ajust de mínims quadrats a partir dels observables (equació 1.1).

$$Z(x, y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2 + b_9y^3$$

Equació 1.1: Superfície de tendència cúbica, polinomi en (x, y) de grau 3

1.3.2 Invers ponderat de la distància

Mètode d'interpolació exacte, local/global i on el valor predit correspon a una mitjana ponderada dels valors observables. El pes assignat a cada mostra depèn de forma inversament proporcional a la distància que el separa del punt problema. El grau en què es modula aquesta dependència és l'exponent i es pot ajustar en funció del model escollit.

La seva fórmula respon a l'equació 1.2:

$$Z(x, y) = \frac{\sum_{i=1}^n \frac{z_i}{r_i^\beta}}{\sum_{i=1}^n \frac{1}{r_i^\beta}} \quad r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

Equació 1.2: Invers de la distància ponderada per l'exponent β que modula la distància r

És un mètode de reduïdes necessitats de càlcul, que depèn tant del nombre de d'observables com de punts problema, senzill d'implementar i que no genera sortides de rang, o sigui, cap valor del model predit està per sobre o per sota dels valors observats.

Existeixen variants locals on en cada punt problema o bé es defineix un radi màxim fora del qual la ponderació es nul·la o bé només participen un nombre determinats d'observables més propers. De fet, aquestes variants locals són habitualment més usades que la metodologia global.

1.3.3 *Splines*

Existeix una família variada de solucions *splines* (p.ex. equació 1.3), però majoritàriament, es pot considerar un mètode d'interpolació quasi-local amb propòsits globals i amb la possibilitat de ser exacte o en cas contrari amb tolerància d'error controlable. Això vol dir que en alguns casos concrets d'aquestes funcions existeix un paràmetre que representa la màxima diferència permesa entre el valor observat i predit.

Els seus models es caracteritzen per la seves formes suaus sense canvis sobtats de valors. Aquest fet pot produir artefactes en zones pròximes amb diferent ritme de variabilitat i també sortides de rang poc mesurades. La seva càrrega de càlcul es pot considerar moderada, depèn bàsicament del nombre d'observables i no del nombre de punts problema, per tant si el nombre d'observables és elevat podria necessitar uns elevats recursos de càlcul.

$$z(x, y) = T(x, y) + \sum_{i=1}^n \lambda_i \cdot R(r_i)$$

$$\rightarrow R(r_i) = - \left\{ \ln \left[\frac{\varphi \cdot r_i}{2} \right] + E_1 \cdot \left[\frac{\varphi \cdot r_i}{2} \right] + C_E \right\}$$

Equació 1.3: *Splines* proposades per [MIT99]

1.3.4 *Kriging*

Mètode d'interpolació geostatístic de difícil classificació ja que en funció del variograma aplicat i de les possibles implementacions simplificades pot ser local, global o sovint un cas intermig amb solucions regionalitzades.

També pot ser un interpolador exacte o no en funció del tipus de solució adoptada. En qualsevol cas és un mètode de complexa implementació, amb certes dificultats en l'aprenentatge i utilització com a usuari i amb una càrrega força important de càlcul [OSU02], [LLO06] i [KER98] que s'analitzaran en el Capítol 4.

En aquest capítol només s'apunten aquests trets més característics, el Capítol 2 està íntegrament dedicat al mètode d'interpolació referència d'aquest treball.

Capítol 2

Kriging

2.1 Introducció

Una primera definició de *kriging* per enginyers especialistes en la geostatística es troba a [DEU98]:

Kriging is a collection of generalized linear regression techniques for minimizing an estimation variance defined from a a prior model for a covariance"

Una segona definició ampliada amb les seves característiques principals realitzada per enginyers del Computer Science Department, University of Southern California apareix a [SHA06]:

Kriging is a complicated interpolation technique developed in the field of geostatistics. The technique observes the underlying process in the space using representative variables (e.g., temperature) and computes unknown values of the variable using the values sampled in a limited set of locations. The interpolation method in Kriging is an optimization procedure which uses a model of the process to determine unknown values. This model is given as a variogram of the process. The method assigns optimal weights to the known values in order to predict the unknown values.

Finalment, una descripció més sintètica en base a la classificació exposada en el Capítol 1, podria ser:

La interpolació mitjançant mètodes geostatístics s'anomena *kriging*.

El terme *kriging* prové de Danie G. Krige, Enginyer de mines i Professor de la Universitat de Witwatersrand a Johannesburg (Republica de Sudàfrica) i la

seva referència inicial és *A statistical approach to some basic mine valuation problems on the Witwatersrand* [KRI51]

Poc després Georges Matheron, també Enginyer de mines, fundador i director del *Centre de Géostatistique et de Morphologie Mathématique* va iniciar la construcció d'uns sòlids fonaments a la seva base teòrica, la Geostatística i el desenvolupament de la Teoria de les Variables Regionalitzades. De Matheron dues són les referències claus: *Traité de géostatistique appliquée* [MAT62] com a obra embrionària de la Geostatística i *The Theory of Regionalised Variables and its Applications* [MAT71], la seva obra més completa.

Kriging va néixer en l'àmbit de l'Enginyeria de mines, on segueix aplicant-se en molts treballs de recerca i també aplicacions tecnològiques, però no es va quedar restringit en aquest àmbit i les seves metodologies es va anar propagant en pràcticament tots els àmbits de l'anàlisi espacial: geologia, estadística, meteorologia, geografia, climatologia, teledetecció, topografia, ecologia, etc.

Per tant, sembla adequat introduir la base teòrica del *Kriging*, la geostatística i posteriorment analitzar les seves principals característiques.

2.2 Geostatística

La **Geostatística** està constituïda pel conjunt de mètodes que estudien la distribució espacial d'una variable des d'una aproximació estadística, estimen els corresponents valors a localitzacions problema o simulen la seva variabilitat.

En els processos d'estimació de la variable es considera que la seva variabilitat espacial no pot ser modelada del tot per una funció matemàtica contínua i suau, i en canvi una part essencial només pot ser descrita per una funció probabilística aleatòria anomenada superfície estocàstica. Aquesta base probabilística constitueix la principal diferència respecte dels habituals mètodes introduïts en el Capítol 1, essencialment deterministes. En conseqüència, el model de variació espacial ha de ser descrit principalment per les dues components anteriorment enunciades [LLO06], però per completar millor el model és necessari afegir una tercera component d'error residual [BUR98] tal com mostra l'equació 2.1. Aquesta component residual, de la mateixa forma que la superfície estocàstica, té una naturalesa aleatòria (no-determinista) però en canvi, no presenta cap tipus d'estructura de correlació espacial.

$$Z(x, y) = M(x, y) + R(x, y) + E(x, y)$$

Equació 2.1: Les tres components representades: *M* tendència, *R* aleatòria estructural i *E* error *r*.

Un variable que pugui ser descrita d'aquesta forma s'anomena **Variable Regionalitzada** [SAM96] i un dels principals objectius de la seva teoria és

modelitzar les dues components que presenten estructura de correlació espacial.

Per a comprendre amb un exemple simple el significat d'una estructura regionalitzada es poden comparar les dues sèries de la figura 2.1 que presenten valors amb les mateixes propietats estadístiques (mitjana, variància, freqüències, etc.) , però diferent estructura espacial:

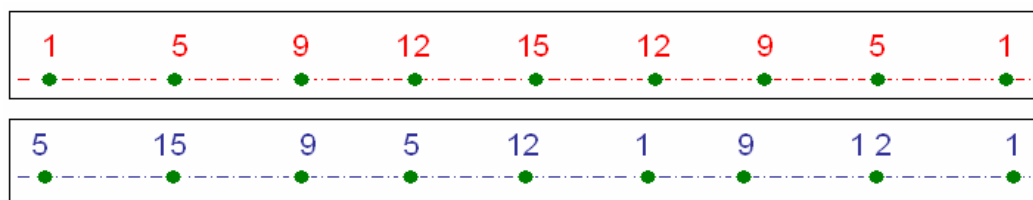


Figura 2.1: Exemple simplificat on s'observa una sèrie (vermella) que correspon a una variable regionalitzada i una sèrie (blava) amb els mateixos valors però sense estructura espacial.

La Teoria de la Variable Regionalitzada (TVR) formula les següents dues hipòtesis sobre les funcions i components expressades en l'equació 2.1:

- Hipòtesi estacionària:
 - El valor esperat de la mitjana és el mateix en totes les localitzacions (equació 2.2)
 - La funció de covariància entre dues posicions separades una distància h només depèn d'aquesta separació h , i no de les posicions absolutes (equació 2.3).
- Hipòtesi intrínseca: El valor esperat de les diferències al quadrat entre dues posicions només depèn de les seves posicions relatives, de la seva separació h , però no de les posicions absolutes (equació 2.4).

$$E\{Z(x)\} = M(x) = m \quad \forall x$$

Equació 2.2: Estacionarietat de la mitjana.

$$C(x+h, x) = E[Z(x+h) \cdot Z(x)] - m^2 = C(h)$$

Equació 2.3: Estacionarietat de la covariància.

$$\gamma(h) = \frac{1}{2} \cdot E\left[\{Z(x+h) - Z(x)\}^2\right] \quad \forall h$$

Equació 2.4: Formulació de la hipòtesi intrínseca de la TVR

Cal observar que en les equacions 2.2, 2.3 i 2.4 (i les posteriors 2.5 i 2.6), x és una simplificació que representa la component espacial en qualsevol

dimensió, en el context d'aquest treball: (X, Y). El formalisme matemàtic d'aquestes expressions està àmpliament desenvolupat en [CRE93] i [KIT97].

Aquestes dues hipòtesis de la TVR tenen conseqüències pràctiques en el procediment d'interpolació *kriging* i constituiran un objectiu a intentar assolir per part de la mostra d'observables i així obtenir els òptims beneficis de la metodologia d'interpolació geostatística. Amb l'objectiu d'acostar-se a la hipòtesi estacionària, caldrà eliminar qualsevol tipus de tendència en la mostra i procurar que sigui el màxim homogènia possible. La hipòtesi intrínseca dóna lloc a l'eina operativa bàsica d'aquest mètode d'interpolació: la generació i modelització del variograma, equació 2.4, que s'analitza en el proper apartat.

2.3 Variograma

Del variograma definit segons l'equació 2.4 es deriva una important conseqüència: la variància de la variable (de fet es pot considerar només la seva component aleatòria, si es verifica la hipòtesi d'estacionarietat) en un context de TVR no depèn de la localització, és una funció de la distància i es dedueix una relació directa entre la variància i el variograma, equació 2.5 que permet gràcies a la hipòtesi intrínseca construir el variograma segons la formulació de l'equació 2.6.

$$\text{Var}[Z(x+h) - Z(x)] = 2\gamma$$

Equació 2.5: Relació entre el variograma i la variància d'una VR

$$\gamma = \frac{1}{2N} \sum_{i=1}^N [Z(x) - Z(x+h)]^2$$

Equació 2.6: Formulació per la generació del Variograma a partir de la mostra d'observables. N correspon al nombre de localitzacions separades una distància h.

Cal indicar que el factor 1/2 de l'equació 2.6 és el que dóna lloc a la denominació de semivariograma per a referir-se fonamentalment al càlcul pràctic d'aquestes diferències al quadrat entre valors a determinades separacions, i a la corresponent representació gràfica, encara que en moltes ocasions s'usen indistintament el dos noms sense cap diferenciació de matís.

La construcció del variograma empíric, a partir dels càlculs indicats en l'equació 2.6 en subdividir el rang de distàncies a analitzar en un nombre determinat d'interval·ls, és un pas indispensable per la interpolació *kriging* i és també una eina bàsica d'anàlisi del patró espacial d'una variable, es comporti o no, com a variable regionalitzada.

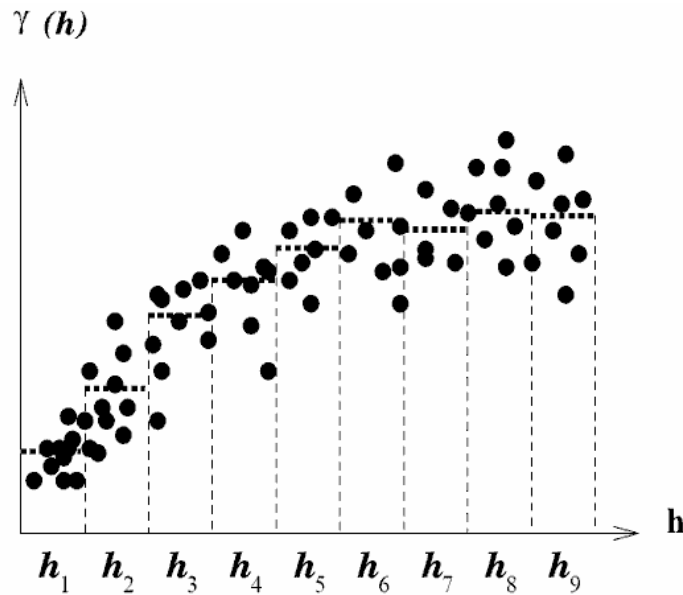


Figura 2.2: Primer pas de la generació del variograma empíric: càlcul i representació de la diferència quadràtica de parells de dades en funció de la seva separació h .

La modelització del variograma empíric per una funció o superposició de funcions contínues és una tasca que té com a objectiu establir els paràmetres estructurals que permeten caracteritzar el patró espacial de la variable i és un pas imprescindible en la interpolació *kriging*. En el Capítol 3 es descriurà l'habitual procés interactiu i la proposta automatitzada d'aquest treball en la generació del variograma modelitzat. En aquest punt, per a comprendre el seu paper dins el procés d'interpolació, la figura 3.5 mostra la diferència entre el variograma empíric i el modelitzat. El variograma empíric s'ha construït segons l'equació 2.6 i la figura 2.4 mostra els elements estructurals: *nugget*, *range* i *sill*, que corresponen als paràmetres a fixar d'un variograma patró.

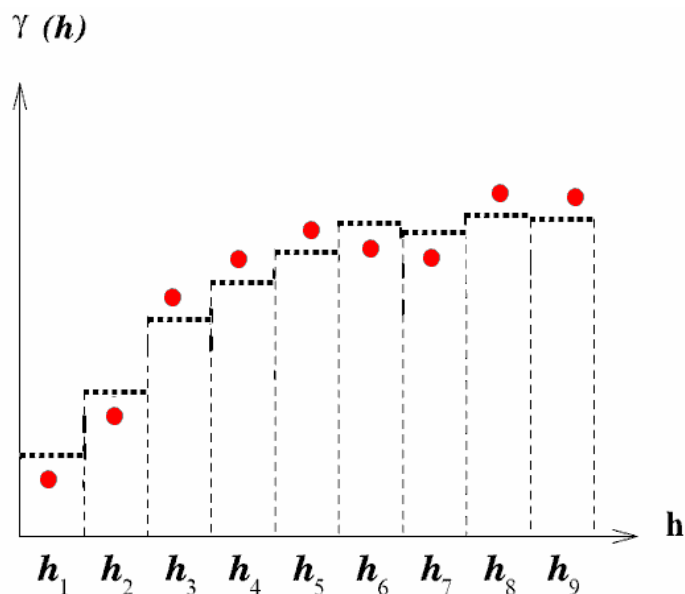


Figura 2.3: Segon pas de la generació del variograma empíric: càlcul i representació de la diferència quadràtica promitjada agrupada per intervals de separacions h

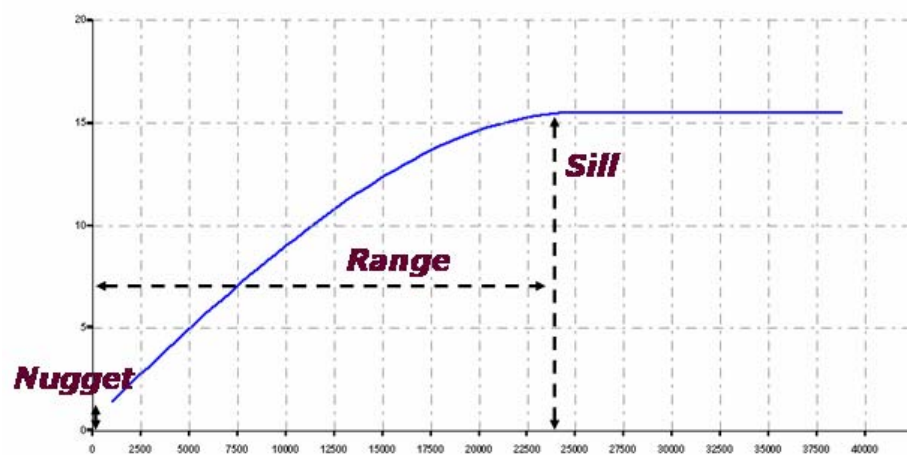


Figura 2.4: Paràmetres estructurals d'un variograma teòric

Unes breus definicions d'aquests elements estructurals són:

- **Nugget:** Desplaçament en l'eix Y que representa la variabilitat en distàncies molt properes a zero i, per tant és un indicador de la component de l'error espacialment no-correlacionat.
- **Range:** Distància on s'assoleix la saturació del variograma, o sigui el patró espacial deixa de presentar un increment de la variabilitat en funció de la distància i s'estabilitza.
- **Sill:** Variància de saturació del variograma. Cal indicar que en models compostos es diferencia el *Partial Sill* com a component de la saturació sense *nugget* del *Sill*, com a variància total (amb *nugget* inclòs).
- **Slope:** En el cas particular del model lineal correspon al pendent de la recta que representa la raó de variabilitat respecte la distància.

Aquests conceptes poden ampliar-se amb descripcions de [BON94] i les valoracions de [OSU02].

2.4. Variants de *Kriging*

La denominació *kriging* s'usa tant per a referir-se de forma global o no específica al conjunt de les diferents metodologies d'interpolació geostatística com pel mètode bàsic específicament anomenat ***kriging ordinari***. Aquest treball també usa aquest doble significat de *kriging*, per exemple en la descripció dels fonaments teòrics s'està usant el significat global i en canvi el títol fa referència a la implementació paral·lelitzada del *kriging* ordinari i no de la resta de variants. A continuació es descriuen breument les variants principals:

- **Kriging simple:** *Kriging* on s'assumeix una mitjana coneguda que es verificarà en el model interpolat. [LLO06] mostra detalls de la formulació i propietats d'aquest mètode.
- **Kriging universal:** *Kriging* superposat a una superfície de tendència (funció polinòmica descrita en el Capítol 1). Aquesta composició de procediments es propaga per tot el procés de forma acoblada i en conseqüència constitueix una metodologia diferenciada respecte a la concatenació seqüencial d'un ajust d'una superfície de tendència i un *kriging* ordinari dels residus (correspon a un dels possibles tipus de *kriging* residual que es descriurà a continuació). Ambdós metodologies tenen per objectiu procurar de convertir en **VR**, una variable poc homogènia que no verifica la hipòtesi d'estacionarietat. A [KIT96] i [CRE93] es poden consultar més detalls de la propietats i les noves expressions de les equacions anteriors amb les funcions acoblades.
- **Cokriging:** Opció multivariant on s'afegeixen un conjunt de variables auxiliars correlacionades amb la variable a modelitzar i que permeten complementar la variabilitat del model de forma més precisa i ajustada. A diferència d'altres mètodes multivariants com el *kriging* residual, *cokriging* permet treure rendiment de la característica habitual que la mostra dels valors observables de les variables auxiliars és molt més nombrosa que la variable principal, com en imatges de teledetecció de diferents resolucions [CHI98] o models multivariants de cerca de determinades propietats del sòl [SAM96]. Sobre els canvis en la formulació es poden revisar els detalls a [PAR07], en particular és interessant observar com es converteix l'equació 2.6 en un variograma creuat entre la variable principal i una variable auxiliar, com mostra l'equació 2.7:

$$\gamma_{AB} = \frac{1}{2N} \sum_{i=1}^N [Z(x)_A - Z(x+h)_A]^2 \cdot [Z(x)_B - Z(x+h)_B]^2$$

Equació 2.7: Formulació del variograma dos dimensional, variables A i B, necessari per la metodologia d'interpolació *cokriging*.

- **Kriging residual:** *Kriging* dels residus (diferències entre els valors observats i els valors modelitzats) obtinguts d'un ajust previ. Aquest ajust previ generador dels residus, té fonamentalment dos orígens diferenciats:
 - Univariant: Pot provenir d'un ajust d'una superfície de tendència que té per objectiu eliminar o reduir inhomogeneïtats en la distribució espacial de la variable, i d'aquesta manera, facilitar al patró real la verificació de la condició d'estacionarietat. Aquesta varietat té per nom *Kriging with external drift* KED i és interessant la referència comparativa [HEN03].
 - Multivariant: El segon tipus d'escenari correspon a un anàlisi multivariant on prèviament s'ha calculat el model de

dependència de la variable d'estudi (independent) respecte possibles variables explicatives (dependents) d'una part important de la seva variabilitat. Aquest model de regressió genera uns residus on s'aplica un procés addicional d'interpolació mitjançant *kriging* ordinari. El procediment combinat està detallat a [NIN00] en un context multivariant amb altres mètodes d'interpolació (deterministes) per a variables climàtiques.

- ***Kriging factorial***: Metodologia multivariant amb variables poc o gens correlacionades. En determinades situacions és necessari un procés previ de descomposició de la variable independent en les seves components principals, segons l'Anàlisi de Components Principals (PCA) [MAN94], per a continuació realitzar diverses interpolacions desacoblades amb *kriging* ordinari de totes o una part de les components obtingudes amb PCA. Més detalls d'aquesta variant es poden consultar a [GOO92].
- ***Fixed Rank Kriging***: Complexa metodologia orientada a reduir l'elevat cost computacional del *kriging*, característica que s'analitzarà a fons en el Capítol 4, quan es disposa d'un nombre elevat de dades. L'estratègia de simplificació de temps (a canvi d'una major complexitat en les formulacions matemàtiques) basada en unes redefinicions de les matrius de covariància es detalla a [CRE08].

Aquestes diferents metodologies exposades no representen una llista exhaustiva de totes les variants de *kriging*, s'han descartat per exemple algunes metodologies amb variables no quantitatives o de suport no puntual (per blocs) per constituir mètodes allunyats del context d'estudi d'aquest treball. En canvi s'ha cregut necessari conèixer, encara que sigui de forma superficial, variants diferents al *kriging* ordinari per a contextualitzar i ampliar la metodologia implementada en aquest treball i apuntar possibles implementacions futures on s'incorporin també els procediments d'ajust i paral·lelització desenvolupats en aquest treball.

2.5. *Kriging* ordinari

En aquest apartat es detalla l'algorisme bàsic del *kriging* ordinari amb l'objectiu principal de donar els elements necessaris perquè en el Capítol 4 es pugui analitzar en quins procediments es consumeix la major càrrega computacional i determinar quin és el disseny adequat de paral·lelització.

Kriging estima el valor de la variable en una localització problema a partir d'una mitjana ponderada dels valors coneguts, tal com es representa en l'equació 2.8 [OLI90]:

$$Z(\vec{x}_0) = \sum_{i=1}^n \lambda_i \cdot Z(\vec{x}_i)$$

Equació 2.8: Valor predit en la localització problema \vec{x}_0 com a mitjana ponderada dels n valors observables en les posicions \vec{x}_i .

L'estratègia de com determinar aquests pesos deriva de la minimització de la variància de la variable tractada com a **VR** a partir de la modelització del seu variograma. Aquesta condició de mínima variància dóna lloc a dues expressions claus que permeten determinar els pesos associats a cada observable (equació 2.9) i que a partir de l'equació 2.8 s'obté el valor predit, i una estimació de l'error associat en la predicció. (equació 2.10) [BUR98]:

$$\sum_{i=1}^n \lambda_i \cdot \gamma(\vec{x}_i, \vec{x}_j) + \phi = \gamma(\vec{x}_j, \vec{x}_0) \quad \forall j$$

Equació 2.9: Equació per la determinació dels pesos que ponderen els observables de la equació 2.8. a partir del variograma modelitzat γ .

$$\sigma_0^2 = \sum_{i=1}^n \lambda_i \cdot \gamma(\vec{x}_i, \vec{x}_0) + \phi$$

Equació 2.10: Equació d'estimació de l'error a partir del variograma modelitzat γ

L'equació 2.9 pot adoptar una notació matricial que serà la més fàcilment traslladable al corresponent codi en una implementació computacional [OSU02].

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \phi \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix}$$

$$A \cdot \lambda = b \rightarrow \lambda = A^{-1} \cdot b$$

Equació 2.11: Formulació matricial de l'equació 2.10 i per àlgebra matricial deducció del vector de pesos λ .

Cal explicar en detall com es construeixen aquestes matrius, per adaptar el seu disseny a un flux parel·lelitzable:

1. Es calculen les distàncies de separació entre tots els parells d'observables

2. Es determina el valor del variograma modelitzat per a cada distància del punt 1.
3. Es construeix la matriu A de l'equació 2.10 amb els valors obtinguts en el punt 2 amb l'ordre mostrat en l'equació 2.11, omplint amb 1 i 0 les posicions que es mostren en les darreres fila/columna.
4. S'inverteix la matriu A .
5. Es calculen les distàncies entre el punt problema i els observables.
6. Es determina el valor del variograma modelitzat per a cada distància del punt 5.
7. Es construeix el vector b (matriu unidimensional) de l'equació 2.10 amb els valors obtinguts en el punt 6 amb l'ordre mostrat en l'equació 2.11, omplint amb 1 la darrera fila.
8. Es realitza el producte matricial de la matriu inversa A^{-1} obtinguda en el punt 4 per el vector b i s'obté el vector de pesos λ que inclou el paràmetre ϕ .

Per a una execució optimitzada de l'algorisme és important tenir en compte que els passos 1 a 4 només cal realitzar-los una vegada ja que només depenen dels observables, que no és variable, excepte pel cas de veïnatge mòbil que s'explicarà més endavant.

Els passos 5 a 8 es realitzen tantes vegades com punts problema es considerin. Quan el resultat de la interpolació és un model digital del terreny (MDT), els punts problema constitueixen el conjunt de centres de cel·la del model ràster resultat, (figura 2.5) i per tant serà la part del procediment on tindrà força sentit la distribució de la càrrega computacional entre diversos processadors.

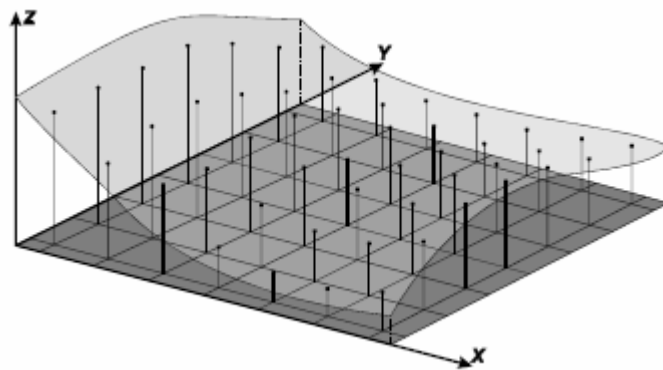


Figura 2.5: Estructura regular d'un MDT on cada centre de cel·la de l'àmbit d'estudi constitueix un punt problema.

[BUR98], [KIT97] i [OSU02] mostren tres exemples numèrics diferents de dimensions senzilles, molt útils per a seguir pas per pas l'algorisme i realitzar verificacions en una implementació computacional.

De forma molt similar a l'obtenció del vector de pesos s'obté el model d'errors per a cada punt problema i per tant no és necessari reproduir el mateix procediment en detall.

Cal remarcar però, que l'estimació de l'error és un dels èxits principals que se l'atribueixen al mètode *kriging* [OLI90].

Even more important, none of the methods provides any estimates of the errors of estimation.

La mateixa referència [OLI90] insisteix a les conclusions:

Using kriging for interpolation enables us to estimate the confidence in any interpolated Value in a way that none of the earlier methods do, and our examples show this.

Però l'estimació de l'error és també una de les principals fonts de conflicte. En el fet que el propi mètode d'interpolació proporcioni un estimador de la qualitat del model generat no hi ha cap controvèrsia i es valora sempre de forma positiva. Ara bé, seria desitjable reafirmar aquest paràmetre de qualitat del model amb un test independent format per dades que no participin en el procediment d'interpolació [PAL95]. El fet que altres mètodes d'interpolació no proporcionin cap estimació de l'error, pràcticament obliga a reservar una submostra dels observables com a test de qualitat, i en canvi, en usar *kriging* molts autors ja no realitzen aquesta reserva confiant totalment en l'estimació de l'error segons l'equació 2.10. Aquesta possible desmesurada confiança en el propi mètode és la que ha creat algunes divergències. En moltes ocasions està justificada per l'elevada complexitat i/o cost del procediment de mostreig i per tant s'optimitza el nombre de mostres necessàries, però en alguns treballs el test independent seria totalment aconsellable i no representaria cap cost addicional i està causat per la inèrcia adquirida en l'ús del mètode o per les facilitats que en l'estimació de l'error proporciona.

Capítol 3

Ajust automatitzat del variograma

3.1 Introducció

Com s'ha mostrat en el capítol 2, la interpolació *kriging* necessita de la modelització del variograma. Aquesta modelització la realitza habitualment l'usuari de forma interactiva, per dos motius:

- Un motiu instrumental: La implementació en la gran majoria de *softwares* consta d'una interfície en mode gràfic que ofereix una sèrie d'eines d'anàlisi del variograma on l'usuari ajudat per la informació geostatística que ofereixen aquestes eines decideix escollir el model i paràmetres concrets del variograma.
- Un motiu analític: El mecanisme interactiu d'exploració de les dades i de determinació del model que millor ajusta, obliga a l'usuari a analitzar el patró espacial de la variable observada i adquirir una millor comprensió del problema que es planteja.

Com l'objectiu principal d'aquest treball és donar una solució completa de paral·lelització del procediment d'interpolació global, cal eliminar qualsevol interacció a mig procés amb l'usuari, i per tant, cal substituir l'habitual cerca interactiva del variograma modelitzat per una solució automatitzada. Cal recordar però, que l'automatització de l'ajust substitueix les raons instrumentals i és la solució que cal aplicar com a pas previ a un procediment de paral·lelització, però en un context més analític caldria que l'usuari usés les eines interactives per a comprendre a fons la distribució espacial de la variable.

3.2 Ajust interactiu del variograma

En aquest apartat es mostra el procediment habitual interactiu descrit per [JIA96] com:

The general practice involves the estimation of the experimental semivariogram for a few lags, followed by manual fitting by trial-and-error of a model selected by the user, usually based on the visual appearance of the experimental semivariogram.

D'aquesta manera és possible avaluar la solució automatitzada que el treball presenta respecte la solució aplicada en la gran majoria de *softwares*. Els detalls de l'ajust interactiu són dependents del *software* concret, però les diferents variants aplicades no són gaire divergents i per tant té sentit exemplificar aquest procediment amb algun *software* concret, tenint en compte que les principals eines gràfiques d'anàlisi del variograma són comunes a tots ells i, en tot cas si cal, s'especificaran algunes de les diferències més significatives.

Els *softwares* analitzats són: **Idrisi**, **MiraMon** i **Surfer**. Idrisi [EAS87] i MiraMon [PON00] [SAN04] són, programaris de Sistemes d'Informació Geogràfica (SIG o GIS) on els observables corresponen a una capa vectorial d'entitats puntuals i la variable a interpolar està, o bé recollida en una base de dades, o forma part d'un atribut addicional. En canvi Surfer [SMI85] pot definir-se com un programari de representació i anàlisi de superfícies.

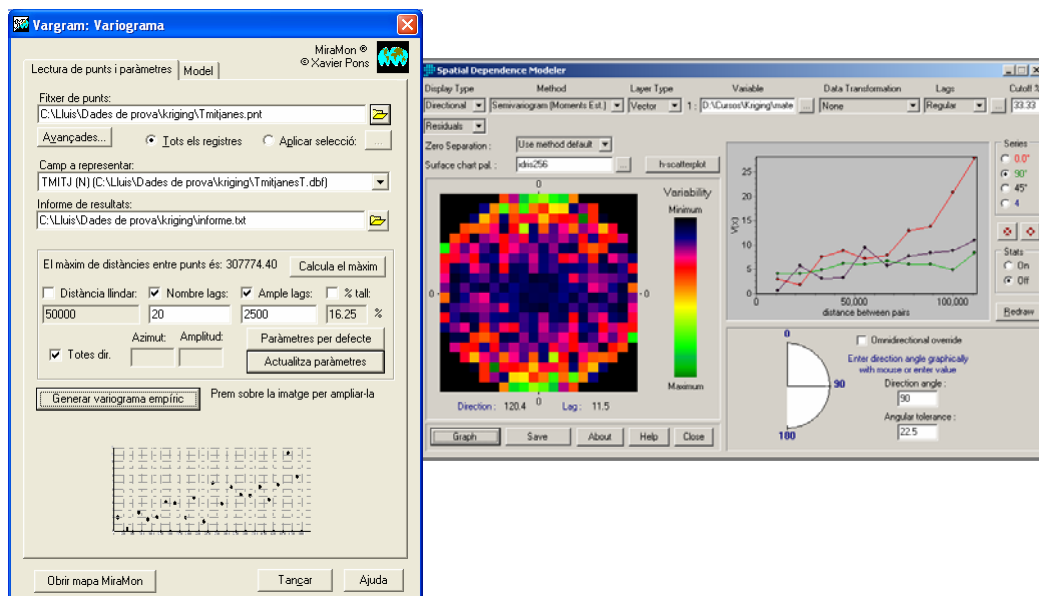


Figura 3.1: A l'esquerra interfície d'anàlisi del variograma del SIG MiraMon. A la dreta interfície equivalent del SIG Idrisi.

3.2.1 Construcció del variograma empíric

En un anàlisi tipus, un usuari estàndard realitzaria els següents passos (amb una interfície de l'estil de la que es mostra en la figura 3.1):

1. Selecció de la mostra d'observables.
2. Si la mostra conté diverses variables, elecció de la variable a modelitzar.

3. Selecció dels paràmetres geomètrics del variograma: amplada de cada interval i nombre d'interval·ls en que es divideix el rang de distàncies a analitzar (a la figura 2.2 es poden identificar aquests paràmetres) .
4. Si es desitja analitzar possible anisotropia, diferent patró espacial de la variable en diferents direccions, selecció de l'angle i amplada angular.
5. Anàlisi visual de la gràfica del variograma en funció dels paràmetres seleccionats amb l'objectiu principal d'identificar de forma aproximada els paràmetres estructurals del variograma (figura 2.4)
6. Anàlisi dels resultats estadístics del variograma, com per exemple distribució de freqüències i distància mitjana a cada interval.
7. Introducció de noves modificacions sobre els paràmetres del punt 3. fins a assolir un variograma estable i amb una estructura propera a la desitjable (figura 2.4)

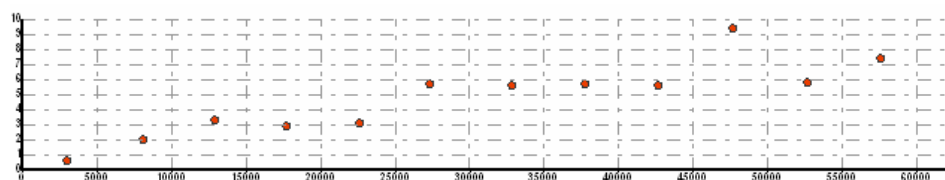


Figura 3.2: Variograma empíric amb 12 intervals d'amplada 5000 m.

En finalitzar aquest procés interactiu s'ha obtingut el variograma empíric (figura 3.2).

3.2.2 Generació del variograma modelitzat

En aquest punt del procediment, cal trobar una funció que s'adapti el millor possible als punts del variograma empíric d'entre la família de possibilitats que ofereix el *software*. Quasi tots els programaris inclouen petites variants de la família de funcions mostrades en la figura 3.3, proporcionades en la seva majoria per [CRE93].

Esquemàticament aquests serien els passos que realitzaria l'usuari:

1. Selecció de l'efecte *nugget* (definit en el capítol 3).
2. Selecció de la funció que visualment s'adapta millor al variograma empíric.
3. Determinació d'uns paràmetres aproximats de la funció escollida, aquests paràmetres, definits en el capítol 3, són els identificatius de l'estructura perseguida: *slope* pel model lineal, *range* i *sill* per la resta
4. Comparació visual entre els variograma model determinats pels paràmetres escollits i el variograma empíric.
5. Iteració del punt 4 fins a una solució considerada apropiada. Si els paràmetres escollits són propers a la solució òptima alguns softwares (p. ex. Idrisi i Surfer) proporcionen eines per acabar d'ajustar-los, però la convergència d'aquest procés no és sempre possible.
6. Seleccionar una funció diferent i repetir els passos 4 i 5.

7. Seleccionar entre les diferents solucions possibles aquella que es consideri adequada. Els programaris proporcionen eines d'anàlisi de l'error entre el variograma empíric i el modelitzat que proporcionen criteris objectius durant aquesta selecció.

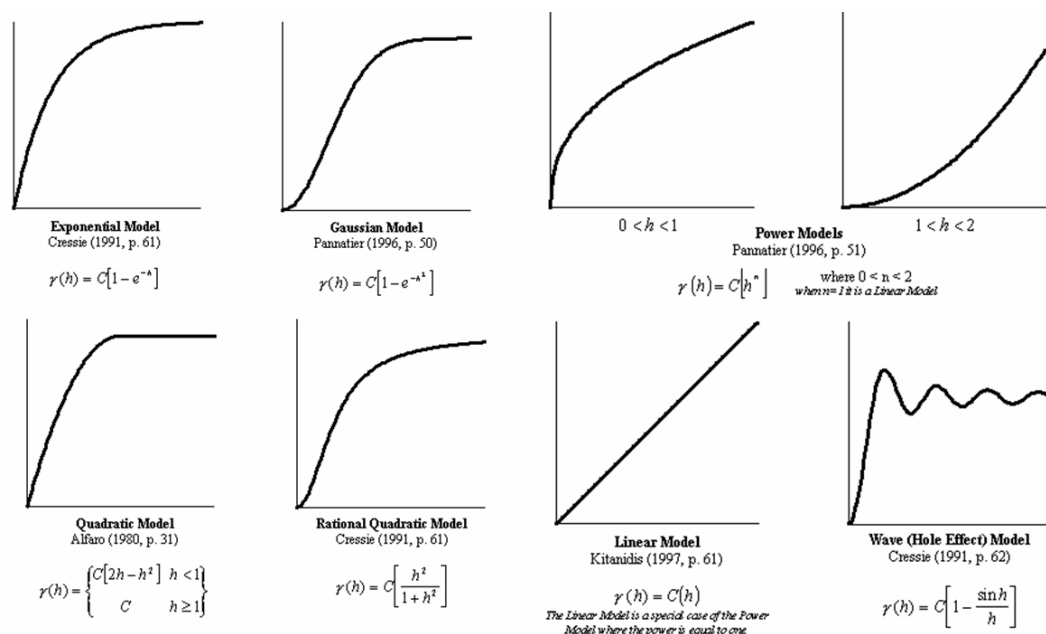


Figura 3.3: Variogrames model de Surfer

Per a realitzar aquests passos, l'usuari acostuma a ajudar-se d'un *software* amb unes eines gràfiques i interactives com les que es mostren a la figura 3.4.

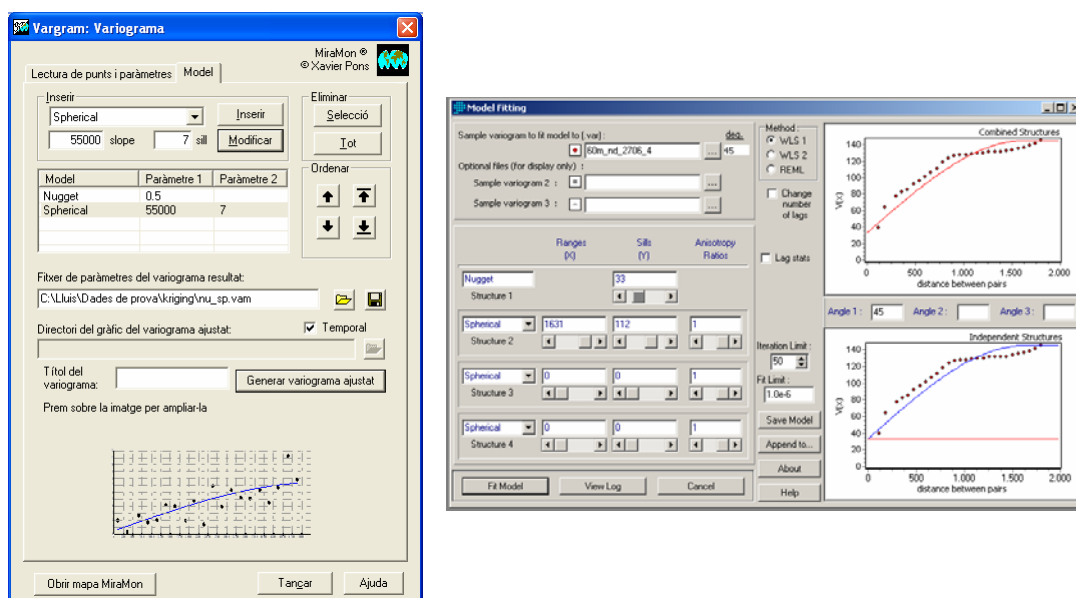


Figura 3.4: A l'esquerra interfície d'ajust del model de variograma del SIG MiraMon. A la dreta interfície equivalent del SIG Idrisi.

En finalitzar aquest procés interactiu s'ha obtingut el variograma modelitzat (figura 3.5).

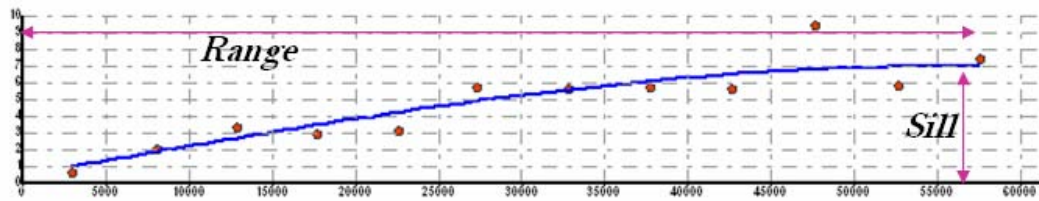


Figura 3.5: Variograma modelitzat i identificació de la seva estructura.

En funció de les dades a analitzar, pot succeir que no sigui possible arribar a l'escenari de la figura 3.5 amb un model ajustat amb una certa estructura. En la bibliografia es recullen diverses propostes per a realitzar un complet anàlisi estructural [SAM96] i així millorar el patró espacial de la mostra o corregir patrons molt allunyats mitjançant transformacions prèvies de la variable [CRE93]. També és aconsellable analitzar el grau de robustesa del resultat aconseguït [SAM96].

3.3 Solució automatitzada per l'ajust del variograma

Diverses són les possibilitats d'ajust d'un variograma concret [CRE93] i [KIT97], o sigui un cas determinat de la família de funcions de la figura 3.3; opció força útil que ofereixen alguns *softwares* però amb determinades dificultats pràctiques, i a vegades, amb poques garanties de convergència de l'ajust. En canvi, són molt poques les referències a un ajust que impliqui les diverses possibilitats alhora, i que, comparant i analitzant les diverses solucions, resolgui quina és la opció més apropiada.

L'article de referència per la solució proposada en aquest treball és [JIA96] on considera que els mètodes d'ajust per mínims quadrats (LS) són clarament més eficients que els mètodes per màxima versemblança (ML). El manual d'usuari del SIG Idrisi [EAS87] també coincideix en qualificar els mètodes ML d'extremadament lents.

La metodologia de cerca de la solució global més òptima que s'ha adoptat en aquest treball està inspirada en l'article de referència [JIA96] i consta dels següents passos:

1. Ajust de forma separada de la solució òptima per a cadascuna de les funcions implementades que són en realitat funcions amb dues components:
 - a. Constant anomenat efecte *nugget*.
 - b. Funció de llistat de possibilitats implementades que s'especifiquen en l'Apèndix.
2. Recollida dels criteris estadístics de cadascun dels ajustos.
3. Comparació per un mateix criteri estadístic dels diferents ajustos i tria de l'ajust més òptim, o sigui el que minimitza l'error d'ajust entre el model empíric i el model ajustat

Diverses són les possibilitats d'ajust per mínims quadrats del punt 1 per a funcions majoritàriament no-lineals com són les de la figura 3.3 i de l'Apèndix.

Analitzant les diferents variants proposades per [MAD04] en aquest treball s'ha optat per el mètode de Levenberg-Marquardt (L-M), exposat en la mateixa referència [MAD04] i de forma original a [LEV44] i [MAR63]. Per la codificació s'ha adaptat la proposta de [PRE88] al context propi de l'ajust de diferents variogrames.

El mètode Levenberg-Marquardt és un procediment iteratiu, on a partir d'uns valors inicials es calculen nous valors per increments temptatius que van ajustant-se als adequats, a partir de les derivades parcials (Apèndix) de la funció respecte els paràmetres a ajustar, i que van convergint, si és possible, fins a la solució final.

En la taula 3.1 s'exemplifica un ajust concret (model esfèric amb *nugget*) d'aquest mètode, segons la solució adoptada en aquest treball, especificant els nous paràmetres a cada iteració. Un resum d'aquesta evolució (només es representen les iteracions 1, 3,5, i 7) es pot analitzar gràficament a la figura 3.6.

Iteració	<i>Nugget</i>	<i>Range</i>	<i>Sill</i>
1	0.07	40252.34	6.78
2	0.28	49223.43	6.34
3	0.25	52904.80	6.88
4	0.49	58621.89	6.83
5	0.47	58422.86	6.92
6	0.48	58650.98	6.89
7	0.41	58689.64	7.01

Taula 3.1: Anàlisi de l'evolució de l'ajust dels paràmetres pel mètode iteratiu L-M.

Cal observar que les dades corresponen al exemple del Capítol 5 i és en aquest capítol del treball on s'exposen les característiques d'aquesta mostra. En l'actual capítol no es considera rellevant conèixer detalls de la mostra, la seva finalitat és simplement il·lustrar amb un exemple el procediment automatitzat de l'ajust.

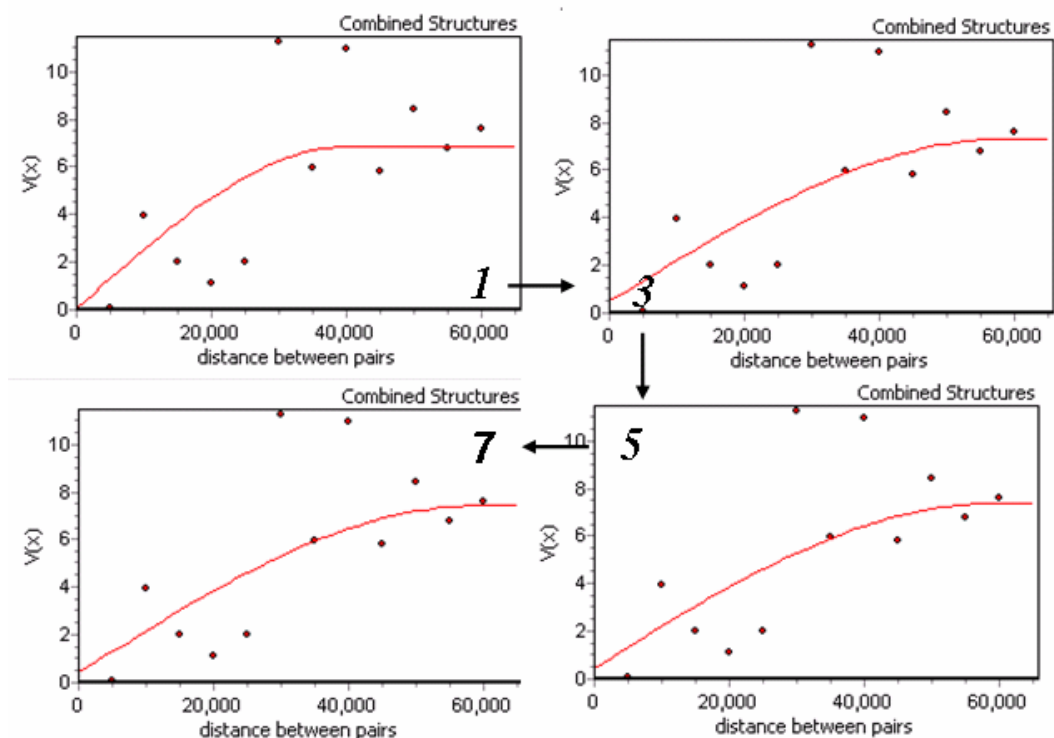


Figura 3.6: Evolució de l'ajust des dels valors inicials: iteració 1, fins a la solució ajustada iteració 7, passant pels valors entremitjos: iteracions 3 i 5. Els paràmetres són els corresponents de la Taula 3.1.

El criteri d'inicialització dels paràmetres ha estat el següent:

- *Nugget*: 10% del valor que correspon al primer interval (el de menor distància) en el variograma
- *Range*: 70% de la distància on es troba l'últim interval
- *Sill*: 90% del valor que correspon a l'últim interval.
- *Slope*: Pendent d'una recta que passaria per un *range* i *sill* com els anteriors o sigui amb una X del 70% i una Y del 90% respecte l'últim interval.

En els diferents exemples s'ha demostrat que aquestes criteris malgrat ser força més simples que els de l'article de referència [JIA96] porten igualment a la convergència de l'ajust. Uns criteris més complexes poden portar més ràpidament a la solució final de l'ajust, però cal observar que el treball de [JIA96] finalitza amb l'ajust i en aquest treball continua amb la interpolació i que, tal i com es mostrarà en el Capítol 4, el temps de càlcul de l'ajust és quasi negligible respecte la interpolació i per tant és en la interpolació on cal posar el màxim esforç per a reduir la càrrega global de còmput.

En la solució proposada s'han recollit 3 estadístics diferents: χ^2 (Chi quadrat), R^2 i R^2 ajustat (que corregeix la dependència de R^2 respecte el nombre de paràmetres) que es mostren en les equacions 3.1, 3.2 i 3.3). Aquesta elecció està inspirada en [PES07] en un context regressió multivariant on es comparen les diferents combinacions de variables

independents possibles per a triar la més òptima [CHA91]. En la proposta d'aquest treball s'ha eliminat algun criteri que no tenia sentit en l'actual context (C_p Mallows) i s'ha afegit el paràmetre χ^2 , donat que és el que usa el mètode L-M per a determinar si en una nova iteració una variació del joc de paràmetres comporta millores en l'ajust o no, i pot ser útil també per a comparar diferents ajustos.

$$\chi^2 = \sum_{i=1}^n \left[\frac{(y_i - f_i)^2}{\sigma_i^2} \right]$$

Equació 3.1: Chi quadrat com a sumatori de les diferències al quadrat entre el valors observats y_i els corresponents a l'ajust f_i amb la possibilitat de ponderar-les amb σ

$$R^2 = \sum_{i=1}^n \left[\frac{(f_i - \bar{y})^2}{(y_i - \bar{y})^2} \right]$$

Equació 3.2: Coeficient de determinació R^2 amb \bar{y} com a mitjana de la mostra i y_i i f_i amb el mateix significat que l'equació 3.1

$$R_{aj}^2 = 1 - \left[\frac{(1 - R^2) \cdot (n - 1)}{n - k - 1} \right]$$

Equació 3.3: Criteri R^2 ajustat, n es nombre de mostres i k el nombre de paràmetres: i R^2 el coeficient de determinació de l'equació 3.2.

En les proves realitzades i que es detallen en el Capítol 4, s'ha verificat que els 3 criteris porten a la mateixa tria, però com no són estrictament equivalents, cal deixar una opció per l'elecció per a resoldre escenaris complexos o molt anivellats de resoldre.

Capítol 4

Paral·lelització de *kriging*

4.1 Justificació

Dels mètodes revisats en el Capítol 1, *kriging* és amb diferència el de major elevat cost computacional. Així ho confirmen múltiples referències com per exemple:

- Des de l'àmbit de l'anàlisi espacial, la primera característica del mètode que [LLO06] destaca és precisament:

Kriging is computationally intensive.

- Des d'un àmbit de computació d'altres prestacions, HPC (*high performance computing*), i per una aplicació de modelització de dades meteorològiques s'afirma:

The Kriging problem is an interesting example of a computationally intensive processing component in a geographic information system (GIS). [KER98]

Però si s'analitza quina és la magnitud d'aquesta diferència, són poques les referències que donen paràmetres quantitatius concrets. Per exemple [PEB97] indica que una mostra de més de 750 punts ja necessita un temps d'execució prou important, però no especifica quina és la dimensió del resultat (l'equació 2.11 mostra que la càrrega de càlcul depèn tant del volum d'observables com de la dimensió del resultat) ni quantifica les mesures de temps. Amb aquest objectiu s'han realitzat algunes proves representatives que confirmen les valoracions referides en la bibliografia i aporten paràmetres mesurats que no sempre es troben de forma completa en les referències.

S'ha realitzat les diferents interpolacions que es mostren a la taula 4.1 a partir d'una mostra de 100 punts generant un resultat de 2673 columnes x 2595 files.

	Tendència	IDW	Splines	Kriging
<i>Idrisi</i>	6	390	No implementat	6042
<i>MiraMon</i>	4	60	147	1347
<i>Surfer</i>	19	53	No implementat	1467

Taula 4.1: Temps d'execució en segons dels 4 mètodes d'interpolació descrits en el Capítol 1 amb els mòduls corresponents de 3 softwares referits en el Capítol 3.

Cal remarcar que les proves no tenen per objectiu realitzar un test dels *softwares*, (caldrà un anàlisi molt més complet), s'han realitzat amb els tres *softwares* diferents (els tres que serveixen per il·lustrar diferents exemples durant aquest treball), amb la finalitat de verificar que la necessitat de còmput del *kriging* comparada amb altres mètodes no és depenent d'una possible implementació particular que pogués donar lloc a conclusions esbiaixades, com confirmen els resultats de la taula 4.1 *kriging* és computacionalment molt més exigent que els altres mètodes analitzats.

Finalment és significatiu remarcar que malgrat aquest important inconvenient pràctic alguns autors manifesten igualment les seves preferències per aquest mètode:

Dubrule (1983, 1984) compared splines and kriging and concluded that although kriging cost more in computing time it led to quantitatively better results. [OLI90]

En canvi altres autors com [SHA06] decideixen que la càrrega de còmput i la complexitat són excessives:

As Kriging is a computationally intensive method, we did not include it in our prototype implementation

En conseqüència sembla justificat realitzar un esforç en reduir de forma significativa el temps de còmput del mètode. Abans d'explicar la solució proposada en aquest treball per paral·lelització, és necessari analitzar la proposta simplificada àmpliament implementada en els *softwares* i usada en múltiples treballs: *kriging with moving neighborhood*.

4.2 Solució aproximada per reducció local

Aquesta solució consisteix fonamentalment en reduir de forma força significativa les dimensions de les matrius (equació 2.11) que s'usen en

l'obtenció dels pesos que ponderen el valor predit a cada punt problema [HES01].

Aquesta reducció de les dimensions i per tant de l'elevat còmput matricial, s'aconsegueix reduint la mostra en un veïnatge proper al punt problema, com per cada punt problema el seu entorn més proper defineix una submostra particular, per aquest motiu aquesta metodologia s'anomena: *kriging with moving neighborhood* [KIT97].

Però aquesta simplificació del problema comporta alhora contradiccions en la seva base teòrica i problemes pràctics que cal com a mínim considerar:

- El problema principal des del punt de vista metodològic radica en que per a realitzar la interpolació s'ha modelitzat i ajustat prèviament un variograma per a unes determinades distàncies que són habitualment força superiors al veïnatge mòbil que s'aplica en aquesta metodologia. En conseqüència, si no es té una cura especial (que cap dels *softwares* analitzats té eines de verificació que ajudin a mantenir la coherència entre el rang del variograma ajustat i el màxim veïnatge, ni tampoc s'ha trobat en la bibliografia que autors que usen aquests *softwares* confirmen explícitament que en les aplicacions pràctiques han vigilat aquesta coherència) amb aquesta metodologia s'està usant només una zona del variograma, quan, en canvi, s'ha ajustat per un àmbit global, àmbit pel qual s'ha pogut identificar la seva estructura espacial.
- En segon lloc cal afegir l'efecte d'interpolació zonal amb discontinuïtats que provoca el fet que les submostres arribin a ser diferents en punts problema propers (figura 4.1) . Aquest efecte, s'explica pel fet que un punt problema recull en el seu veïnatge particular una determinada submostra, i per tant, el valor predit serà una suma ponderada dels valors d'aquesta submostra, i és possible que un punt problema corresponent a una cel·la adjacent, que per veïnatge mòbil seleccioni una submostra lleugerament diferent, doni lloc a un valor predit significativament diferent.

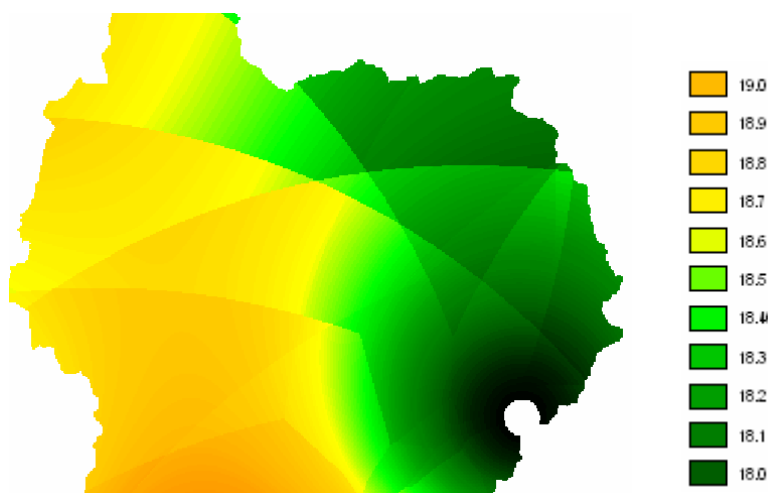


Figura 4.1: Discontinuitats en el model resultat provocades per la metodologia de veïnatge mòbil.

Malgrat els problemes pràctics i teòrics considerats, molts *softwares* implementen aquesta metodologia (figura 4.2) i, en alguns casos, la consideren obligatòria quan detecten un nombre de mostres considerable.

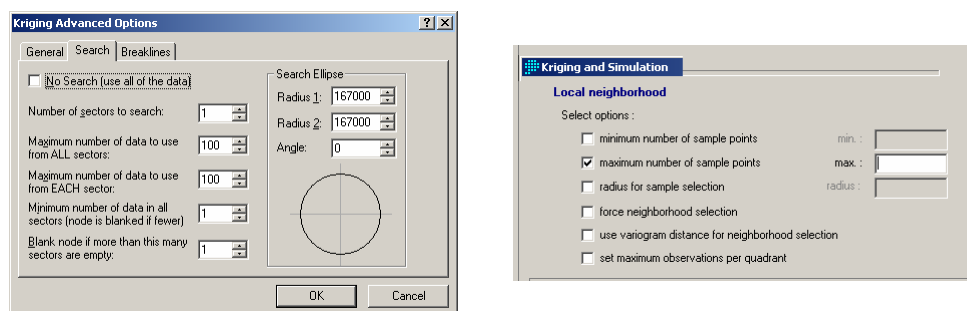


Figura 4.2: Exemples (Surfer interfície esquerra i Idrisi interfície dreta) de configuració dels paràmetres de veïnatge mòbil com a simplificació de *kriging*.

En conseqüència, és necessari constatar que els possibles efectes de la metodologia de veïnatge mòbil no sempre són una solució vàlida com a reducció de l'elevat cost computacional de *kriging*, que quan el temps de càlcul no és assumible en una aplicació concreta, és més aconsellable optar per un dels altres mètodes més lleugers mostrats en el Capítol 1, que realitzar una simplificació d'aquesta metodologia, que pot reduir considerablement la qualitat del model. [KIT96] confirma aquests consells:

In conclusion, using a moving neighborhood is not recommended as a general practice (although it may be useful in special cases) and will not be discussed further in this book.

4.3 Solució completa per paral·lelització

En l'apartat anterior s'ha argumentat la necessitat de mantenir la solució global que interpola el model a partir de la mostra d'observables completa per a obtenir uns valors de predicció de qualitat. Si no és possible disminuir la càrrega global de computació, sembla una solució adequada repartir-la entre diferents processadors. A partir de l'anàlisi del mètode d'interpolació que es detalla en el Capítol 2 (apartat 2.5) s'ha desenvolupat un procediment de paral·lelització del codi mitjançant l'estàndard MPI (Message Passing Interface) [GRO99] i [PAC97] que s'exposa de forma detallada durant aquest apartat.

4.3.1. Solució MPI

MPI és una proposta d'implementació estàndard per la transmissió de missatges entre processadors en un model de computació paral·lelitzada [MPI94] i [SNI96]. En l'actualitat és la proposta majoritària de paral·lelització successora de PVM.

El codi que implementa la solució adoptada en aquest treball és en llenguatge C. Aquest codi s'ha bifurcat amb directives de precompilació (*#ifdef*) per a poder commutar entre una execució paral·lelitzada que inclou les llibreries MPI i una execució en sèrie, que s'ha demostrat força útil per a perseguir errades en la implementació de l'algorisme on no està implicat el procediment de paral·lelització i realitzar proves comparatives de rendiment. S'ha treballat inicialment en un entorn de proves emulat d'un únic processador amb sistema operatiu Windows XP Professional i llibreries MPICH2 per Windows [ASH08] per a compilador MS Visual Studio 6.0. Posteriorment, quan les execucions en l'entorn emulat eren satisfactòries, s'ha traslladat el codi i tot el banc de proves a un entorn distribuït amb sistema operatiu Linux (Debian etch 4.0) i llibreries MPI 1.2.7. L'executable per aquest entorn no-emulat, s'han generat amb el compilador GCC 4.12, a partir del mateix codi, fet que mostra que la solució és portable i funcionalment operativa, i vàlida per a com a mínim, aquests dos sistemes operatius.

Encara que existeixen determinades solucions de paral·lelització de *kriging* (en nombre molt reduït si es compara amb un important nombre d'implementacions sèrie en programaris d'anàlisi espacial, paquets estadístics, Sistemes d'Informació Geogràfica, etc...) no s'ha trobat una solució equivalent, portable en C i amb llibreries MPI. Aquest és un recull dels precedents de paral·lelització de *kriging* trobats:

- [KER98] solució en CMF (*Connection Machine Fortran*), solució exclusiva per a un entorn distribuït molt concret, CM5.
- [GEB03] solució en PVM (*Parallel Virtual Machine*) per al software estadístic R.
- [ROS03] solució acoblada al *software* estadístic R a partir d'un entorn d'alt nivell on el desenvolupador accedeix a l'API del *software* però no directament a la implementació no detallada en PVM o MPI.

4.3.2 Disseny de la paral·lelització

Per a un disseny correcte de la paral·lelització, cal analitzar com és el flux d'una execució sèrie del model. Aquest anàlisi està detallat en el Capítol 2 (apartat 5) i s'observa com hi ha una fase (punts 1 a 4) comú, i una fase (punts 5 a 8) força iterativa, que es repeteix per a cada punt problema. Aquesta reiteració cal abordar-la des del coneixement de l'estructura espacial del conjunt dels punts problemes que en el context d'aquest treball formen un model digital del terreny.

En el Capítol 1 es va introduir el concepte de model digital del terreny que s'il·lustra en el la figura 2.5. Aquest model conté de forma implícita una estructura regular on l'àmbit d'estudi, o sigui la regió global de tots els punts problema del model, està distribuït en un conjunt de files i columnes formades per una unitat fonamental que és la cel·la (figura 4.3). Cada cel·la dóna lloc a un punt problema i cal considerar si la solució és enviar la resolució del model per a cada punt problema a un node de còmput, o

s'aprofita la seva organització espacial per a agrupar de forma coherent un conjunt de càlculs i estalviar un possible excessiu trànsit d'enviament de procediments i dades.

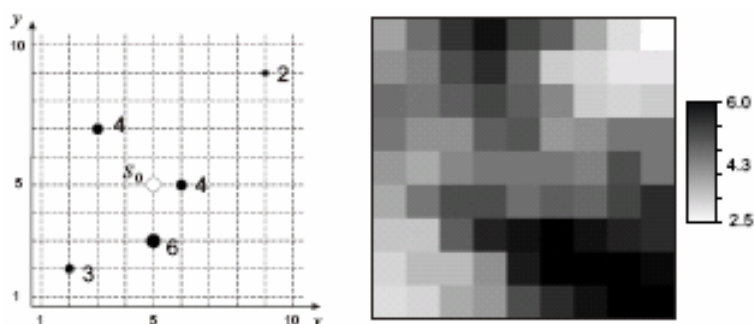


Figura 4.3: Mostra sobre l'estructura espacial del model resultat i exemple de resultat en una representació tessel·lada.

La solució que s'ha adoptat consisteix en construir una unitat mínima de distribució de dades: la fila de punts problema.

El disseny per aquesta unitat fila implica que el nucli de generació del model d'interpolació és un operador de files de processament amb els següents paràmetres bàsics (representats esquemàticament a la figura 4.5):

- Rep un identificador de fila.
- Processa el model per a cada cel·la de la fila.
- Torna una fila resultat.

Per a processar l'algorisme del model, el procés (operador) per files necessita prèviament una conjunt d'informació que parametriza el model segons els conjunt de dades (invariant per files) :

- Nombre total de files, columnes i mida de costat de cel·la.
- Coordenades de la regió d'estudi (en el context d'aquest treball referides a un SRH introduït en el capítol 2).
- Model de variograma, ajustat de forma automàtica com es detalla en el capítol 4.
- Dades, valors i posicions dels observables.

De forma complementària a la fila de resultat, l'operador retorna uns valors addicionals:

- el màxim i mínim de la fila de valors; útil per a obtenir el mínim i màxim de tot el model i així donar una correcta informació del rang de valors als *softwares* visualitzadors del model digital resultat.
- l'identificador de la fila per a construir el model digital ordenat
- possible codi d'error per a controlar l'execució.

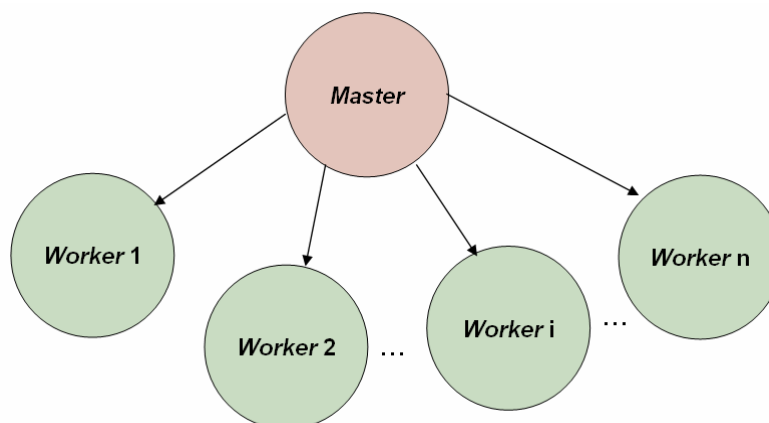


Figura 4.4: Estructura jeràrquica *master/worker*

Analitzat el disseny bàsic de paral·lelització s'opta per un model *master/worker* on cada processador *worker* només es comunica amb el coordinador o *master* (estructura jeràrquica Figura 4.4). La tasca dels *workers* consisteix en realitzar els processament de l'operador de files anteriorment descrit cada vegada que el *master* li envia una fila com s'il·lustra en esquemàticament en la figura 4.5.

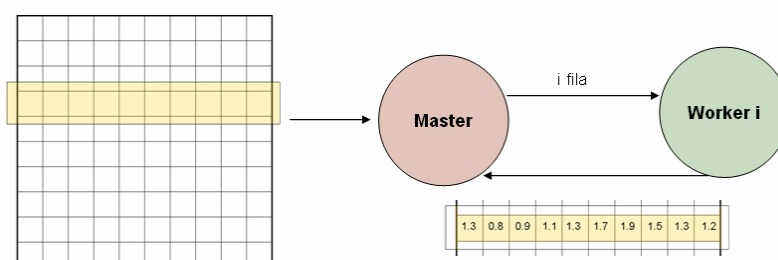


Figura 4.5: Transferència de missatges bàsic (simplificat) *master/worker*

El paper del *master* fonamentalment consisteix en:

- Enviar ordres de processament de files als diferents *workers*
- Recollir els resultats
- Guardar els resultats en un suport no volàtil

Aquestes tasques bàsiques han de requerir especial cura per part del *master* en els següents punts:

- Cal un mínim temps entre la rebuda d'una fila i l'enviament d'una de nova: cal aprofitar al màxim la disponibilitat de processament dels *workers*.
- Cal controlar correctament l'ordre d'enviament i rebuda de les files (que pot ser diferent), el resultat podria no tenir sentit en desfer-se la seva estructura espacial.
- Cal controlar adequadament la finalització progressiva de les feines dels *workers* (l'experiència ha demostrat que aquest ha estat un dels punts més delicats).

A més a més, cal fer notar que el *master* realitza tot els procediment previ que no té sentit paral·lelitzar ja que el temps necessari és negligible, però que cal recordar:

1. Lectura de dades des del format original
2. Construcció del variograma empíric.
3. Ajust dels diferents models de variograma implementats.
4. Elecció del variograma que ajusta millor.

En aquest punt cal afegir l'enviament únic a tots els *workers* de les dades especificades com a invariants. En la implementació en MPI això implica un enviament amb la instrucció **MPI_Bcast()** com es mostra a tall d'exemple en la figura 4.6:

```
#ifndef VERSIO_MPI
MPI_Type_contiguous(3, MPI_DOUBLE, &p_punt);
MPI_Type_commit(&p_punt);
MPI_Type_extent(MPI_INT, &intex);
displacements[0] = (MPI_Aint) 0;
displacements[1] = intex;
MPI_Type_struct(2, blocklengths, displacements, tipus_model, &p_model);
MPI_Type_commit(&p_model);
MPI_Bcast(&n_punts, 1, MPI_INT, 0, MPI_COMM_WORLD);
if (id_proc!=0)
{
    if ((dat=calloc(n_punts, sizeof(*dat)))==NULL)
    {
        puts("ERROR: Sense memoria"); return 1;
    }
    if ((array_inversa_mostres_A=calloc((n_punts+1)*(n_punts+1),
        sizeof(*array_inversa_mostres_A)))==NULL)
    {
        puts("ERROR: Sense memoria"); return 1;
    }
}
MPI_Bcast(&n_col, 1, MPI_INT, 0, MPI_COMM_WORLD);
MPI_Bcast(&costat_pixel, 1, MPI_DOUBLE, 0, MPI_COMM_WORLD);
MPI_Bcast(&x_inici, 1, MPI_DOUBLE, 0, MPI_COMM_WORLD);
MPI_Bcast(&y_inici, 1, MPI_DOUBLE, 0, MPI_COMM_WORLD);
MPI_Bcast(&model_i, 1, p_model, 0, MPI_COMM_WORLD);
MPI_Bcast(&sol, 1, MPI_INT, 0, MPI_COMM_WORLD);
MPI_Bcast(dat, n_punts, p_punt, 0, MPI_COMM_WORLD);
MPI_Bcast(array_inversa_mostres_A, (n_punts+1)*(n_punts+1),
    MPI_DOUBLE, 0, MPI_COMM_WORLD);
```

Figura 4.6: Codi d'exemple de la transferència massiva dels paràmetres invariants a tots els processadors.

Després del procediment distribuït, i una vegada tots els *workers* han finalitzat la seva tasca reiterada d'interpolació de files, el *master* reordena el model resultat perquè tingui l'estructura espacial necessària. L'estratègia d'escriure en un fitxer temporal les dades resultat amb l'ordre d'arribada, optimitza el procediment, ja que el *master* redueix el temps de coordinació i per tant d'espera dels *workers*. Quan ja s'ha acabat la interpolació, és una tasca ràpida reescriure en un nou fitxer el model resultat amb l'ordre preestablert, on per facilitar i verificar l'ordre correcte en model definitiu, el fitxer temporal té guardat un identificador de control de fila.

Capítol 5

Resultats experimentals


5.1 Descripció de l'entorn experimental

Per a validar el correcte disseny i implementació de la solució exposada en el Capítol 4, s'han realitzat dues proves amb característiques força diferenciades en tipus de variable, àmbit d'estudi, escala de detall i nombre de mostres que s'especifiquen en els propers apartats. Aquestes dues validacions també tenen unes característiques comunes que a continuació s'exposen:

Les proves s'han realitzat en el *cluster* del Laboratori d'Arquitectura de Computadors de l'Escola Tècnica Superior d'Enginyeria (ETSE) de la Universitat Autònoma de Barcelona (UAB). Els nodes d'aquest *cluster* són PCs Pentium IV 2,6 GHz amb 768 Mb de RAM que es comuniquen amb una xarxa Ethernet a 100Mbps.

Ambdós casos les dades originals estaven en format MMZ, format comprimit del SIG MiraMon i s'han transformat a format DAT, format ASCII del *software* Surfer on s'expliciten les coordenades X,Y i els valors observables (Figura 5.1 esquerra). Aquest format DAT és el que necessita com a entrada de les variables el model proposat en aquest treball.

Els resultats interpolats que genera el model desenvolupat en aquest treball corresponen a un model de dades ràster en format de text GRD. Aquest format és també un format de text ASCII del *software* Surfer que consta bàsicament d'una capçalera amb unes metadades mínimes i a continuació emmagatzemats seqüencialment els valors de la variable per a cada cel·la, ordenats per files i amb un retorn com a separador de files (figura 5.1 dreta).

<pre> 288065.000000 4596135.000000 23.000000 306483.000000 4625962.000000 22.600000 295930.000000 4605320.000000 22.700001 298500.000000 4634000.000000 22.100000 304742.000000 4630460.000000 22.500000 299800.000000 4509000.000000 23.400000 291480.000000 4563745.000000 23.600000 482863.000000 4662623.000000 22.700001 426880.000000 4585800.000000 22.500000 408696.000000 4571567.000000 20.100000 301350.000000 4548625.000000 24.400000 346410.000000 4589053.000000 22.000000 </pre>		<pre> DSAA 2673 2595 260330 527530 4488810 4748210 9.509511548826 26.099785685562 21.197971349045 21.201036389563 21.204103252797 2: 21.228688748142 21.231766347408 21.234844494893 2: 21.259473376196 21.262550406651 21.265626623438 2: 21.290328280852 21.293452933584 21.296586746989 2: 21.321963789118 21.325171278315 21.328385955924 2: 21.354337347733 21.357607359318 21.360882453213 2: 21.387618769208 21.391114579216 21.394649654605 2: </pre>
--	---	---

DAT

GRD

Figura 5.1: Mostra del format d'entrada (DAT) i del format de sortida (GRD). Ambdós en mode text però en un model de dades diferent, vector i ràster respectivament.

De forma complementària i totalment externa al model, aquest format text GRID s'ha importat al format binari IMG propi de MiraMon amb dos propòsits:

- Realitzar uns mínims tests per validar si la interpolació és suficientment correcta (a partir de la submostra que no participa en la interpolació), però fonamentalment per descartar possibles errades tant en la implementació de l'algorisme com en el procés de reorganització del resultat distribuït. És important recordar que l'esforç prioritari d'aquest treball és reduir el temps de càlcul mantenint qualitat predictiva sense simplificacions.
- Visualitzar, consultar els resultats, donar una simbolització adequada als models generats i aplicar uns límits més adequats, sovint irregulars, als àmbits d'estudi real (el model els genera rectangulars a partir de les coordenades mínimes i màximes de la regió d'anàlisi).

5.2 Interpolació de temperatures

En aquesta primera validació de resultats, la variable a interpolar correspon a la temperatura mitjana mensual de les mitjanes diàries del mes d'agost de 2005. Aquestes dades van ser facilitades pel Servei Meteorològic de Catalunya (SMC) dins un conveni de desenvolupament tecnològic amb el Centre de Recerca Ecològica i Aplicacions Forestals (CREAF) [PES07] i corresponen a valors estadístics de les dades que proporciona la xarxa automàtica d'estacions meteorològiques del propi SMC distribuïdes per tot el territori de Catalunya.

La distribució espacial de les estacions es pot observar en la figura 5.2. De la mostra total es van seleccionar 100 estacions, que corresponen aproximadament al 66% del total i així reservar una submostra per a validar el model resultat.

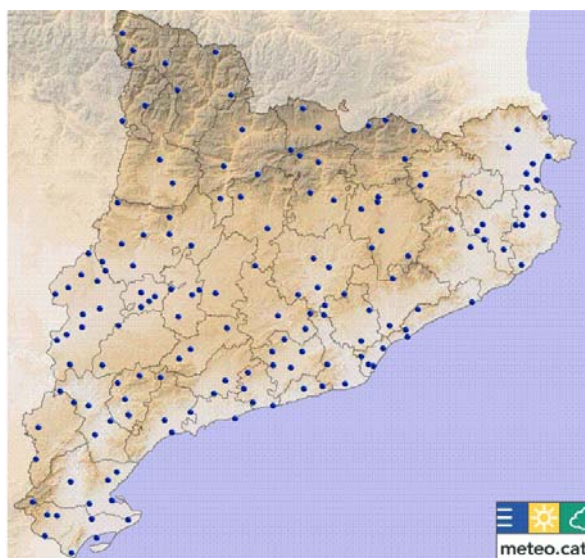


Figura 5.2: Localitzacions de les estacions (punts blaus) de la xarxa automàtica del SMC.

La mostra seleccionada (les 100 estacions, la resta no intervé en cap moment de la generació del model interpolat) es caracteritza per les dades recollides en la taula 5.1 i l'àmbit d'estudi pels paràmetres de la taula 5.2.

Nombre dades	100
Màxim (C)	26.1
Mínim(C)	9.5
Rang (C)	16.6
Mitjana (C)	21.0
Desv. Std (C)	3.5

Taula 5.1: Valors estadístics de la mostra de temperatures.

Superfície total (km ²)	69284.96
Costat de cel·la (m)	100
Nombre columnes	2672
Nombre files	2593
Nombre de cel·les	6928496

Taula 5.2: Paràmetres de la regió d'estudi, àmbit rectangular que conté tot el territori de Catalunya

Una caracterització descriptiva i comparada amb el segon joc de dades del proper apartat que resumeixi les taules 5.1 i 5.2, permet qualificar que aquesta mostra com a reduïda, d'àmbit extens i que amb un resolució espacial moderada genera una matriu de dimensions considerables de punts problema.

El variograma empíric a partir dels paràmetres introduïts com a arguments del mòdul que realitza l'ajust i la interpolació s'observa en la figura 5.3 representat per punts. Aquest variograma dona lloc al millor variograma

ajustat de forma automàtica pel model representat per una línia en la figura 5.3, que en aquest cas és un model esfèric, amb els paràmetres i el valors dels criteris estadístics que qualifiquen l'ajust detallats en la taula 5.3.

<i>Nugget</i>	0.44
<i>Range</i>	58463.64
<i>Sill</i>	6.94
χ^2	11.74
R^2	0.821
R^2 ajustat	0.754

Taula 5.3: Paràmetres del variograma ajustat automàticament i de la qualitat de l'ajust en la interpolació de temperatures.

La representació gràfica del variograma ajustat es mostra també en la figura 5.3. Aquest model de variograma és el que fixa els paràmetres de interpolació *kriging*, tal com s'ha exposat en els Capítols 2 i 3.

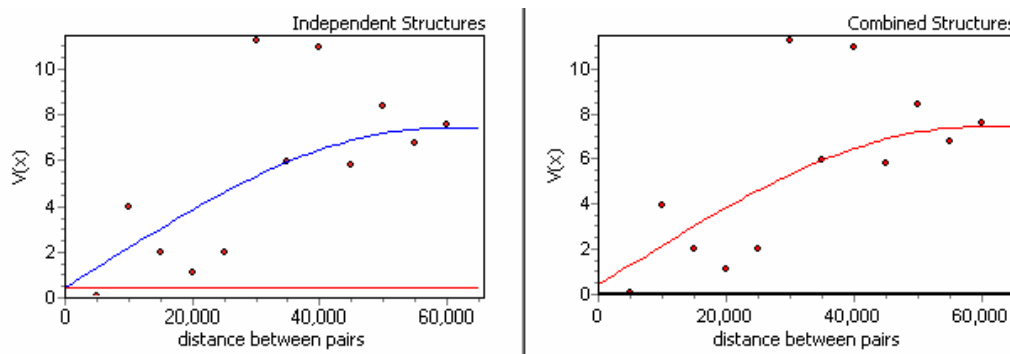


Figura 5.3: Representació gràfica del variograma empíric (punts vermells) i del modelitzat en la interpolació de temperatures. A l'esquerra descompost en *Nugget* (línia vermella) i Esfèric (línia blava) i a la dreta combinat (línia vermella).

A partir d'aquest punt, la càrrega de còmput es distribueix segons el disseny de paral·lelització exposat en el Capítol 4 i els resultats del rendiment es mostren a la taula 5.4 i la corresponent gràfica d'escalabilitat de la figura 5.4.

N Workers	Temps (s)	Temps (mm:ss")
0*	1029.23	17 '09"23
2	414.60	6' 54"60
4	208.14	3' 28"14
8	110.81	1' 58"81
12	87.03	1' 27"03
14	90.41	1 '30"41
16	88.26	1'28"26

Taula 5.4: Taula de rendiment de la paral·lelització per la interpolació de temperatures. (*) En el disseny aplicat *master/worker*, el nombre de processadors correspon al nombre de workers + 1 (el *master*), per tant 0 *workers* correspon a un únic processador en un procediment en sèrie.

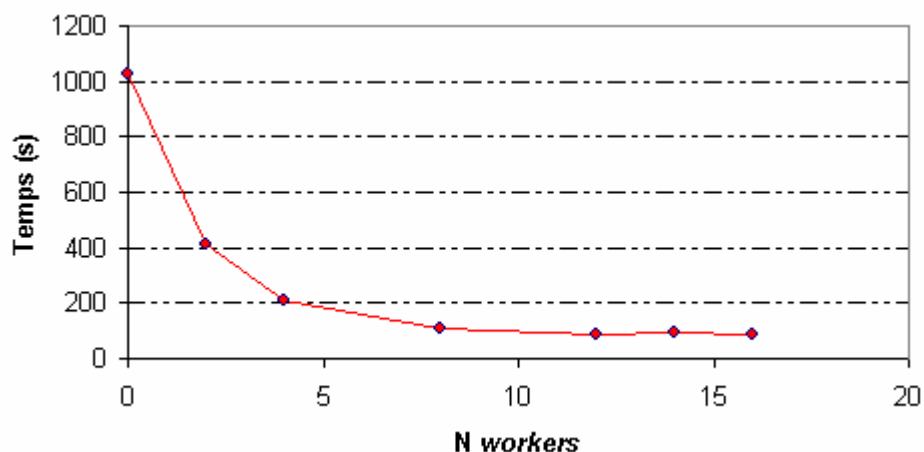


Figura 5.4: Gràfica d'escalabilitat del model d'interpolació de temperatures.

En la gràfica de la figura 5.4 es pot observar una correcta escalabilitat fins arribar a assolir una zona de saturació a partir de 12 *workers*.

Un anàlisi complementari és la determinació del *speed-up*, quocient entre el temps sèrie i el temps corresponent a n processadors totals (comptant el *master*). Aquest anàlisi es veu reflectit en la taula 5.5 i la figura 5.5.

N Processadors	Speed-up
1	1.000
3	2.482
5	4.945
9	9.288
13	11.826
15	11.384
17	11.661

Taula 5.5: Taula d'*speed-up* derivada dels temps de la Taula 5.4.

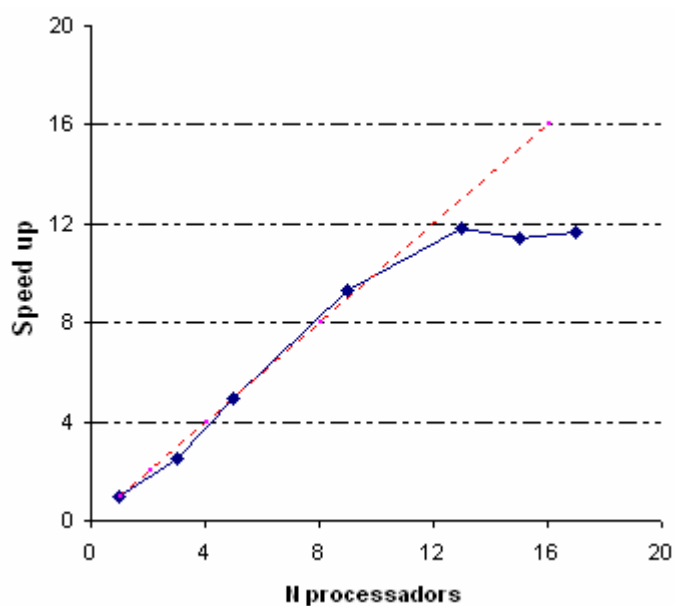


Figura 5.5: Anàlisi d'*speed-up* amb les dades recollides a la Taula 5.5.

Per a completar l'anàlisi de rendiment s'ha afegit una llibreria de creació de fitxers *log* de registre de totes les transmissions entre processadors. Aquest *log* permet, amb el programari *Jumpshot* [CHA07], identificar els possibles problemes de transmissió i aplicar millores en el rendiment. Un exemple d'anàlisi gràfic amb 9 processadors (8 *workers* , 1 *master*) amb *Jumpshot* es mostra a la figura 5.6.

En resum, s'ha aconseguit reduir el temps d'execució d'aproximadament 17 minuts a 1 minut i mig.

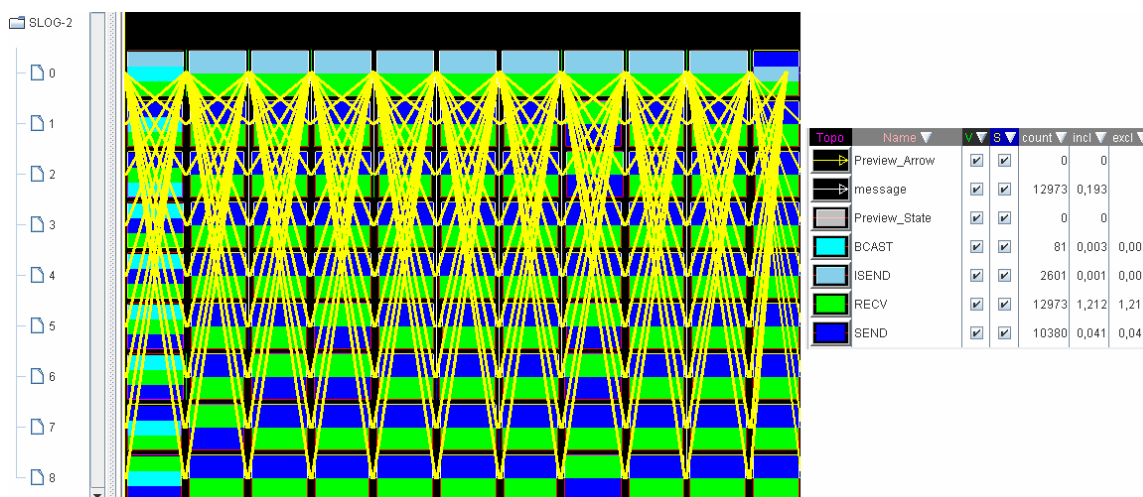


Figura 5.6: Anàlisi gràfic de transmissions entre un *master* i 8 *workers*

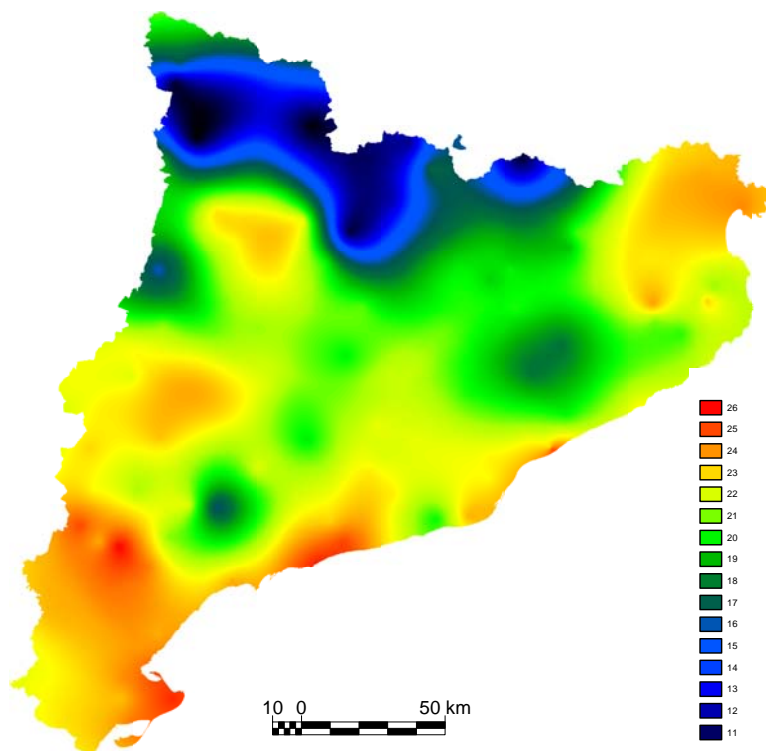


Figura 5.7: Mapa del model interpolat de temperatures .

5.3 Interpolació d'altituds

La segona prova de validació s'ha realitzat amb la variable altitud ortomètrica a partir d'una submostra de les mesures de posició tridimensional realitzades durant dos anys de campanya 2007 i 2008 al Jardí Botànic de Barcelona. Aquestes mesures s'han realitzat amb un receptor GPS model Leica1200, model que incorpora el geoide adequat per a transformar alçades el·lipsoïdals a alçades ortomètriques [MAU07].



Figura 5.8: Vista parcial de la distribució de la mostra amb valors altitudinals.

La distribució espacial de les mesures realitzades amb el receptor GPS és mostra en la figura 5.8, els seus valors estadístics en la taula 5.6 i el seu àmbit d'estudi en la taula 5.7.

Nombre dades	2855
Màxim (m)	137.2
Mínim (m)	101.2
Rang (m)	36.0
Mitjana (m)	115.6
Desv. Std (m)	7.0

Taula 5.6: Valors estadístics de la mostra

Superfície total (m ²)	121600
Costat de cel·la (m)	0.5
Nombre columnes	760
Nombre files	640
Nombre de cel·les	486400

Taula 5.7: Paràmetres de l'àmbit d'estudi localitzat al Jardí botànic de Barcelona (Montjuïc).

En aquest segon cas, es tracta, comparativament a la primera mostra, d'una mostra amb un volum important de dades en un àmbit reduït i una resolució de força detall. Globalment, degut al nombre elevat d'observables, és una prova amb necessitats molt més importants de temps de computació i on s'evidencia la necessitat de distribuir la càrrega de càlcul per aquesta metodologia *kriging*, tal i com es proposa en aquest treball.

L'ajust automatitzat del variograma empíric resol, en aquesta situació, un model sinusoidal amb *nugget* amb els paràmetres i valors estadístics que es mostren exposats en la taula 5.8.

<i>Nugget</i>	-1.19
<i>Range</i>	30.75
<i>Sill</i>	52.78
χ^2	32.173
R^2	0.995
R^2 ajustat	0.993

Taula 5.8: Paràmetres del variograma ajustat automàticament i de la qualitat de l'ajust en la interpolació d'altituds.

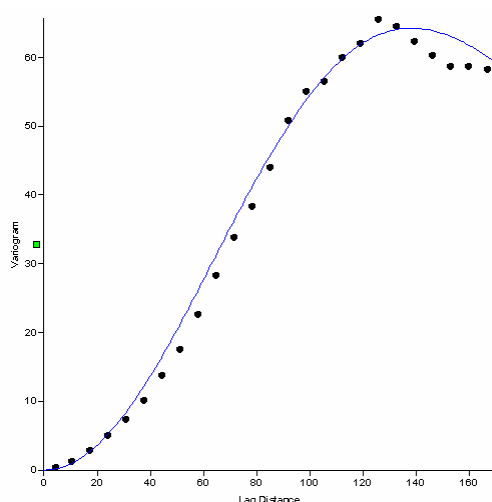


Figura 5.9: Representació gràfica del variograma empíric (punts negres) i del modelitzat (línia blava) corresponent al model sinusoidal.

N Workers	Temps (s)	Temps (hh mm'ss")
0*	41172.00	11h 26' 12"00
2	22559.37	6h 15' 59"37
4	9942.034	2h 45' 42"03
8	5373.00	1h 29' 33"00
12	4196.83	1h 09' 56"83
14	4699.19	1h 18' 19"19

Taula 5.7: Taula de rendiment de la paralelització per la interpolació d'altituds

En la gràfica de la figura 5.10 es pot observar una correcta escalabilitat que arriba fins a 12 processadors. A partir d'aquest punt i sense zona de transició torna a créixer el temps de càlcul. En aquest exemple, s'ha aconseguit reduir el temps d'execució de quasi 11 hores i mitja en una execució en sèrie a quasi 1 hora i 10 minuts en la òptima distribuïda.

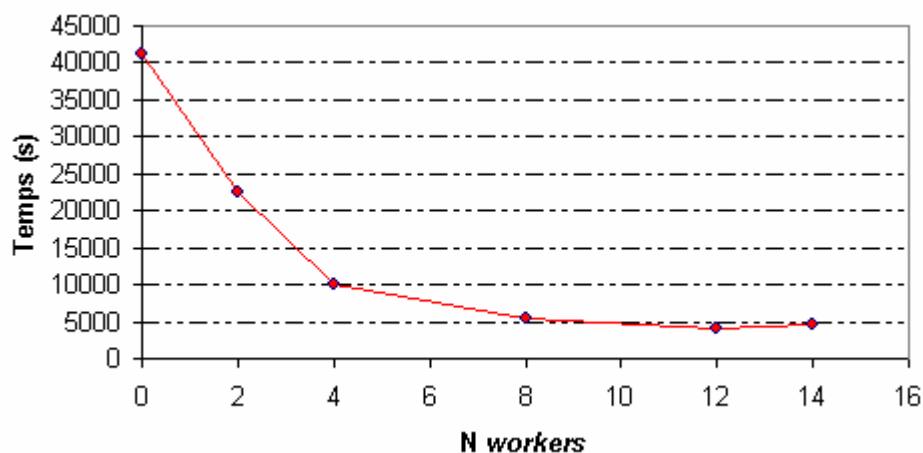


Figura 5.10: Anàlisi d'escalabilitat del model interpolat d'elevacions.

L'anàlisi de rendiment complementari pot observar-se en la figura 5.11 amb els valors d'*speed-up* de la taula 5.8.

N Processadors	<i>Speed-up</i>
1	1.000
3	1.825
5	4.141
9	7.663
13	9.810
15	8.762

Taula 5.8: Taula d'*speed-up* derivada dels temps de la Taula 5.7.

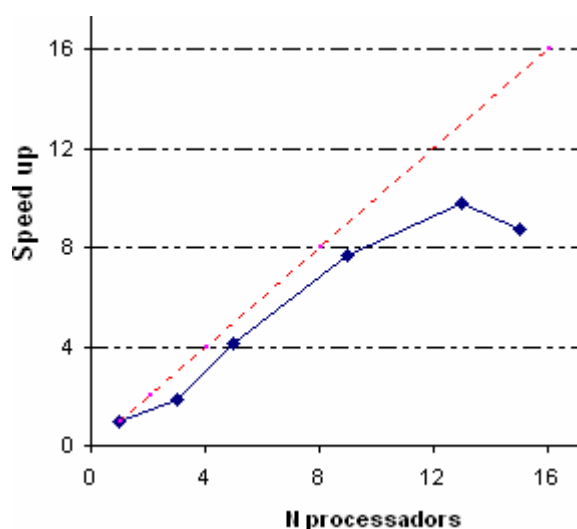


Figura 5.11: Anàlisi d'*speed-up* amb els valors de la Taula 5.8.

Finalment es mostra en la figura 5.12 un mapa d'elevacions del model resultat elaborat amb eines SIG a partir del resultat que genera la interpolació.

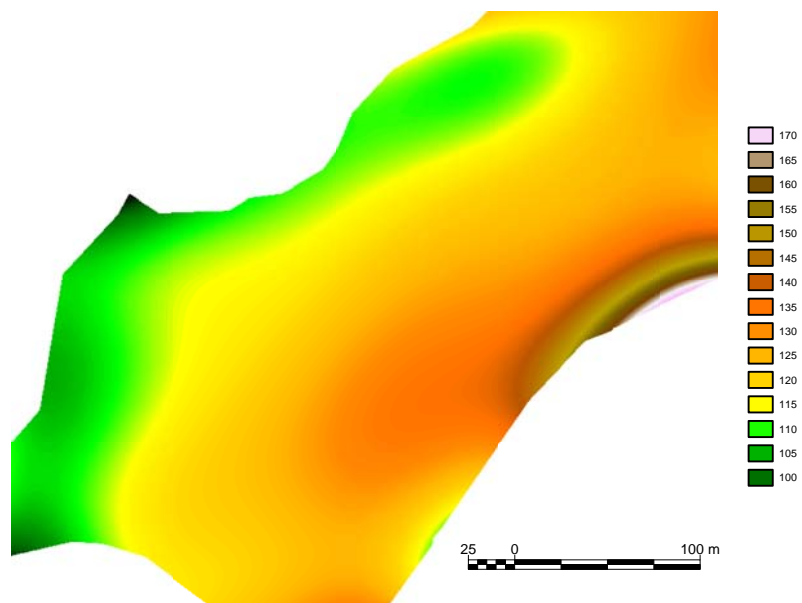


Figura 5.12: Mapa del model interpolat d'elevacions.

Capítol 6

Conclusions i treballs futurs

El treball que es presenta ha situat i caracteritzat el mètode d'interpolació *kriging* dins el conjunt de mètodes d'interpolació i ha mostrat amb referències i validacions l'elevat cost computacional d'aquesta metodologia.

La solució que ofereix aquest treball no va en la línia majoritària de simplificar la metodologia. Aquest treball opta per mantenir la complexitat del mètode i oferir una implementació paral·lelitzada mitjançant MPI sobre llenguatge ANSI C, estàndars que són garantia de portabilitat a diferents plataformes.

Amb l'objectiu d'aconseguir un adequat rendiment en el procés paral·lelitzat s'ha analitzat l'algorisme, de forma especial estudiant amb detall els passos amb major càrrega computacional i quines propietats de les aplicacions d'aquesta metodologia influeixen en aquest còmput.

Per aconseguir l'objectiu d'un procés d'interpolació complet en mode paral·lel ha estat necessari automatitzar l'ajust del variograma i substituir l'habitual procediment interactiu que el determina. La solució d'aquest ajust s'ha fonamentat en implementar el mètode Levenberg-Marquardt, mètode iteratiu específic per ajust de funcions no-lineals, a un conjunt de variogrames model i escollir amb criteris estadístics el model òptim d'entre els ajustats.

Els anàlisis de rendiment en dos exemples de característiques ben diferenciades mostren que la paral·lelització del procés ha aconseguit l'objectiu principal de baixar de forma significativa els temps d'execució sense pèrdua de qualitat en el model interpolat, i per tant, confirmen la viabilitat de la solució proposada.

La proposta d'aquest treball articulada en dos fases, ajust del variograma i paral·lelització dels algorismes d'interpolació *kriging* és traslladable a altres metodologies d'interpolació geostatística que comparteixen la seva complexitat i elevades necessitats de computació, en conseqüència marca una línia de continuació futura en recerca i desenvolupament tecnològic. Concretament representaria una continuació natural la paral·lelització del mètode *cokriging* descrit breument en el Capítol 2.

Una futura línia amb elements de continuació d'aquest treball, però que també obre noves possibilitats, consistiria en aplicar tant les metodologies geostatístiques com les estratègies de distribució de càrrega de computació a l'àmbit de la Teledetecció. Dues referències que caldrà analitzar, on es recull un ampli ventall de possibilitats en aquest sentit són: *Geostatistics and remote sensing* [CUR98] i *High Performance Computing in Remote Sensing* [PLA07].

Bibliografia

- [ASH08] D. Ashton, W. Gropp and E. Lusk *Installation and User's Guide to MPICH, a Portable Implementation of MPI*. Argonne National Laboratory, University of Chicago, 1996, ANL/MCS-TM-ANL-01/96/5.
- [BON94] G.F. Bonham-Carter, *Geographic information systems for geoscientists modelling with GIS*, Pergamon. 1994, p. 398.
- [BUR98] P.A. Burrough and R.A. McDonnell, *Principles of Geographical Information Systems*, Oxford University Press, 1998, p. 333.
- [CHA91] S. Chatterjee, and B. Price. *Regression Analysis by Examples*. 2nd Edition J. Willey & Sons. New York, 1977, p. 278
- [CHA07] A. Chan D. Ashton and R. Lusk *Jumpshot- Users Guide* Argonne National Laboratory, University of Chicago, 2007
- [CHI98] M. Chica-Olmo and F. Abarca-Hernandez "Radiometric coregionalization of Landsat TM and SPOT HRV images", *International Journal of Remote Sensing*, 1998, vol. 19-5, pp. 997-1005.
- [COO04] W. Cooper and C. Jarvis "A Java-base intelligent advisor for selecting a context-appropriate spatial interpolation algorithm" *Computers & GeoSciences* 2004, vol. 30 pp. 1003-1018.
- [CRE93] N.A.C. Cressie *Statistics for Spatial Data*, Wiley Series in Probability and Mathematical Statistics John Wiley & Sons New York 1993, p. 900.
- [CRE08] N.A.C. Cressie and G. Johannesson "Fixed rank kriging for very large spatial data sets" *Journal of the Royal Statistical Society: Series B*, 2008, vol. 70 I.1, pp. 209-226.
- [CUR98] P. J. Curran P. M. Atkinson "Geostatistics and remote sensing" *Progress in Physical Geography*, 1998, vol. 22, No. 1, pp. 61-78.

- [DEU98] C.V. Deutsch and A.G. Journel, *GSLIB- Geostatistical Software Library and User's Guide*. 2nd Edition Applied Geostatistical Series. Oxford University Press, New York, 1998, p. 369.
- [EAS87] J. R. Eastman, *IDRISI Project*. 1987, Clark Labs, Clark University, Worcester, USA. <http://www.clarklabs.org>.
- [FDG98] FGDC Std. 001-1998, *Content Standard for Digital Geospatial Metadata* Federal Geographic Data Committee. Washington 1998.
- [FEL94] Á.M. Felicísimo, *Modelos digitales del terreno. Introducción y aplicaciones en las ciencias ambientales*, Biblioteca de Historia Natural, 3. Pentalfa Ediciones. Oviedo, 1994, p. 220.
- [FRA82] R. Franke, "Scattered data interpolation: tests of some methods", *Mathematics of Computation* 1982, vol. 38, pp.181-199.
- [GEB03] A. Gebhardt, "PVM Kriging with R", *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 2003, Vienna, Austria K. Hornik, F. Leisch & A. Zeileis (eds.), <http://www.ci.tuwien.ac.at/Conferences/DSC-2003>
- [GEO94] GEODERMA. Spatial Interpolation, Elsevier Science B.V. Amsterdam, 1994 Vol 62, pp.17-231
- [GOO92] P. Goovaerts, "Factorial Kriging Analysis: A Useful Tool for Exploring the Structure of Multivariate Spatial Soil Information." *Journal of Soil Science* 1992, vol 43, pp.597-619.
- [GRO99] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI: Portable Parallel Programming with the Message-Passing Interface*, MIT Press, 1999.
- [HEN03] T. Hengl, G. Heuvelink and A. Stein "Comparison of kriging with external drift and regression kriging" Technical note International Institute for Geo-information Science and Earth Observation (ITC), Enschede, 2003, p. 18.
- [HES01] M. Hessami, F. Anctil and A. A. Viau "Delaunay implementation to improve kriging computing efficiency", *Computers & Geosciences*, 2001, vol. 27, pp.237-240
- [ISO01] ISO Std. 19111 *Geographic information – Methodology for feature cataloguing* International Organization for Standardization: Technical Committee 211, 2001
- [JIA96] X Jian, R. A. Olea R.A. and Y. Yu "Semivariogram modeling by weighted least squares" *Computers & Geosciences*, 1996, vol. 22,-4, pp. 387-397
- [KER98] K.E.Kerry and K.A.Hawick *Kriging Interpolation on High-Performance Computers* Technical Report DHPC-035 Department of Computer Science, University of Adelaide, Australia, 1998.

- [KIT97] P.K. Kitanidis, *Introduction to geostatistics: applications to hydrogeology.*, Cambridge University Press. 1997, p. 249
- [KRA06] J.F. Kratzer, D.B. Hayes, and B. E. Thompson, "Methods for interpolating stream width, depth, and current velocity", *Ecological Modelling* 2006, 196 pp.256–264.
- [KRI51] D.G. Krige "A statistical approach to some basic mine valuation problems on the Witwatersrand" *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 1951, vol. 52, 6, pp.119-139.
- [LAU92] R. Laurini, and D. Tompson, *Fundamentals of Spatial Information Systems* Academic Press. Londres 1992, p. 680.
- [LEV44] K. Levenberg, "A method for the solution of certain non-linear problems in least squares" *Quart. Applied Math.*, 1944,, vol 2,-2, pp.164-168.
- [LLO06] C. D. Lloyd *Local Models for Spatial Analysis*, CRC Press, Belfast 2006, p. 244.
- [MAD04] K. Madsen, H.B. Nielsen and O. Tingleff, *Methods for non-linear least squares problems*, Informatics and Mathematical Modelling Technical University of Denmark 2004, p. 58.
- [MAN94] B. F. J. Manly, *Multivariate Statistical Methods.*, Chapman & Hall/CRC Boca Raton, 1994 (2nd edition), p. 215.
- [MAR63] D. E. Marquardt, "An algorithm for least squares estimation of nonlinear parameters": *Journal of the Society for Industrial and Applied Mathematics.*, 1963, vol. 11- 2, pp. 431-441.
- [MAT62] G. Matheron, G. *Traité de Géostatistique appliquée.*, Tome 1, Editions Technip, Paris, 1962, p. 334.
- [MAT71] G. Matheron, *The theory of regionalized variables and its applications*, Fontainebleau,. Ecole de Mines, 1971, p. 212.
- [MAU07] D.F. Maune *Digital Elevation Model Technologies and Applications: The DEM Users Manual* American Society for Photogrammetry and Remote Sensing, Bethesda, 2007, p. 620.
- [MIT93] H. Mitsova and L. Mitso, "Interpolation by Regularized Spline with Tension". *Mathematical Geology.* 1993, vol. 25, 6, pp. 641-655.
- [MPI94] Message Passing Interface Forum. "MPI: A Message-Passing Interface standard. International" *Journal of Supercomputer Applications*, 1994, vol. 8, pp.165–414.
- [NIN00] M. Ninyerola, X. Pons and J.M. Roure, "A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques". *International Journal of Climatology*, 2000, vol., 20: pp. 1823-1841.

- [OLI90] M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information systems", *International Journal of Geographical Information Science*, vol 4-3, pp. 313 - 332
- [OSU02] D. O'Sullivan and D. Unwin, *Geographic Information Analysis*. John Wiley & Sons, Hoboken New Jersey, 2002, p. 436.
- [PAC97] P.S. Pacheco, *Parallel Programming with MPI* Morgan Kaufman Publishers, 1997, p. 418.
- [PAL95] V. Palà and X. Pons, "Incorporation of relief into geometric corrections based on polynomials" *Photogrammetric Engineering & Remote Sensing*, 1995, 61-7, pp. 935-944
- [PAR07] E. Pardo-Igúzquiza and P.M. Atkinson "Modelling the semivariograms and cross-semivariograms required in downscaling cokriging by numerical convolution-deconvolution" *Computers & GeoSciences* 2007, vol. 33, pp.1273-1284.
- [PES07] L. Pesquer, J. Masó and X. Pons "Integración S.I.G. de regresión multivariante, interpolación de residuos y validación para la generación de rásters continuos de variables meteorológicas". *Revista de Teledetección.*, 2007, vol 28, pp. 69-76
- [PLA07]: A. J. Plaza and C.I Chang *High Performance Computing in Remote Sensing* Chapman & Hall/CRC Computer & Information Science, Boca Raton, 2007, p. 466.
- [PON00] X. Pons, *MiraMon. Sistema d'Informació Geogràfica i software de Teledetecció*, Centre de Recerca Ecològica i Aplicacions Forestals, CREAF. Bellaterra, Spain. <http://www.creaf.uab.cat/MiraMon>.
- [PRE88] W. H Press, S. A Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C* Cambridge Univ. New York, 1988, p. 963
- [ROS03] A. Rossini, L. Tierney and N. Li, *Simple Parallel Statistical Computing in R*, UW Biostatistics Working Paper Series University of Washington, 2003, Paper 93, <http://www.bepress.com/uwbiostat/paper193>.
- [SAM96] F.J. Samper and J. Carrera, *Geoestadística. Aplicaciones a la Hidrogeología*, Universitat Politècnica de Catalunya, Barcelona 1996, p. 484.
- [SAN04] J.M. Santos Preciado, *Funcionamiento del programa MiraMon. Aplicación para la realización de ejercicios prácticos de carácter medioambiental y/o territorial. Cuaderno de Prácticas*. UNED. Madrid. 2004, p. 167.
- [SHA06]: M. Sharifzadeh and C. Shahabi, "Utilizing Voronoi Cells of Location Data Streams for Accurate Computation of Aggregate Functions in Sensor Networks" *Geoinformatica* 2006, vol 10-1, pp. 9-36.

- [SMI85] D. Smith and P. Madison: *Surfer. Surface Mapping System* Golden Surfer Inc. Colorado USA. <http://www.goldensoftware.com>.
- [SNI96] M. Snir, S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra The *MPI Complete Reference* Massachusetts Institute of Technology Press, 1996, p. 352.
- [TOB70] W. R. Tobler, "A computer model simulation of urban growth in the Detroit region". *Economic Geography*, 1970, vol. 46, 2, pp. 234-240

Apèndix

Variogrames: Equacions

A continuació es detallen les equacions dels variogrames implementats en el model i les fórmules de les seves derivades parcials respecte els seus paràmetres. Aquestes derivades són necessàries per a l'ajust del variograma segons el mètode de Levenberg-Marquardt descrit en el Capítol 4.

Cal remarcar dues propietats que apliquen a tots els models:

- Tenen definit un paràmetre *nugget* addicional a la seva formulació més elemental.
- Tenen forçada explícitament una discontinuïtat al voltant del zero, per d'aquesta manera assegurar que són un interpolador exacte, tal com justifica [KIT97]. Això vol dir que la fórmula es divideix en dues parts:

$$\gamma(h) = \begin{cases} h=0 & \Rightarrow \gamma=0 \\ h>0 & \Rightarrow \gamma=\gamma(h) \end{cases}$$

També cal observar que en les fórmules, **N** representa al paràmetre *nugget*, **R** és el *range*, i **S** el *sill* i **P** el pendent (*slope*); *h* representa la variable distància i γ és la funció semivariograma. El significat d'aquests paràmetres i variables estan exposats al Capítol 3.

Linial

$$\gamma(h) = N + P \cdot h$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial P} = h$$

Esfèric amb saturació a R

$$\gamma(h) = \begin{cases} h \leq R \Rightarrow \gamma(h) = N + S \cdot \left(\frac{3}{2} \cdot h_r - \frac{1}{2} \cdot h_r^3 \right) \\ h > R \Rightarrow \gamma = N + S \end{cases}$$

$$h_r = h(R) = \frac{h}{R} \Rightarrow \frac{\partial h_r}{\partial R} = -\frac{h}{R^2} = -\frac{h_r}{R}$$

$$h \leq R$$

$$h > R$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial R} = \frac{\partial \gamma}{\partial h_r} \cdot \frac{\partial h_r}{\partial R} = \left(-\frac{3}{2} \cdot \frac{S}{R} \right) \cdot (h_r - h_r^3)$$

$$\frac{\partial \gamma}{\partial N} = 0$$

$$\frac{\partial \gamma}{\partial S} = \frac{3}{2} \cdot h_r - \frac{1}{2} \cdot h_r^3$$

$$\frac{\partial \gamma}{\partial N} = 1$$

Exponencial

$$\gamma(h) = N + S \cdot \left(1 - e^{-h_r} \right)$$

$$h_r = h(R) = \frac{h}{R} \Rightarrow \frac{\partial h_r}{\partial R} = -\frac{h}{R^2} = -\frac{h_r}{R}$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial R} = \frac{\partial \gamma}{\partial h_r} \cdot \frac{\partial h_r}{\partial R} = -\frac{S \cdot h_r \cdot e^{-h_r}}{R}$$

$$\frac{\partial \gamma}{\partial S} = 1 - e^{-h_r}$$

Quadràtic

$$\gamma(h) = N + S \cdot (2 \cdot h_r - h_r^2)$$

$$h_r = h(R) = \frac{h}{R} \Rightarrow \frac{\partial h_r}{\partial R} = -\frac{h}{R^2} = -\frac{h_r}{R}$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial R} = \frac{\partial \gamma}{\partial h_r} \cdot \frac{\partial h_r}{\partial R} = \frac{2 \cdot S \cdot h_r \cdot (h_r - 1)}{R}$$

$$\frac{\partial \gamma}{\partial S} = 2 \cdot h_r - h_r^2$$

Gaussià

$$\gamma(h) = N + S \cdot \left(1 - e^{-\frac{h^2}{r}} \right)$$

$$h_r = h(R) = \frac{h}{R} \Rightarrow \frac{\partial h_r}{\partial R} = -\frac{h}{R^2} = -\frac{h_r}{R}$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial R} = \frac{\partial \gamma}{\partial h_r} \cdot \frac{\partial h_r}{\partial R} = -\frac{2 \cdot S \cdot h_r^2 \cdot e^{-h_r^2}}{R}$$

$$\frac{\partial \gamma}{\partial S} = 1 - e^{-h_r^2}$$

Sinusoidal

$$\gamma(h) = N + S \cdot \left(1 - \frac{\sin h_r}{h_r} \right)$$

$$h_r = h(R) = \frac{h}{R} \Rightarrow \frac{\partial h_r}{\partial R} = -\frac{h}{R^2} = -\frac{h_r}{R}$$

$$\frac{\partial \gamma}{\partial N} = 1$$

$$\frac{\partial \gamma}{\partial R} = \frac{\partial \gamma}{\partial h_r} \cdot \frac{\partial h_r}{\partial R} = \frac{S}{R} \cdot \left(\cos h_r - \frac{\sin h_r}{h_r} \right)$$

$$\frac{\partial \gamma}{\partial S} = \left(1 - \frac{\sin h_r}{h_r} \right)$$